

REMARKS

Claims 21-30 and 32-40 are pending in the application. Claims 32-34 and 38-40 are withdrawn as being drawn to non-elected inventions. Claim 31 has been cancelled. Claims 21-31 and 35-37 are under consideration. Claims 21, 22, 25, 29, 30, and 36 have been amended to further clarify the intended subject matter of the claimed invention. Entry of these amendments is respectfully requested. Applicants reserve the right to prosecute non-elected subject matter in subsequent divisional applications.

Information Disclosure Statement

Although not required for the submission of an information disclosure statement, Applicants have enclosed an alignment of the sequence of SEQ ID NO:12 with the sequences of references 1 and 2 from the IDS filed October 21, 2002 in order to clarify the relevance of the reference sequences (please see CLUSTALW alignment attached as Exhibit A).

Amendments to the specification

Priority

The specification has been amended to add the priority information as suggested by the Examiner in order to comply with 35 U.S.C. § 119(e) and 37 C.F.R. § 1.78. Applicants previously made a proper claim to priority under Article 8 of the Patent Cooperation Treaty (pages 1-2 of the Declaration and Power of Attorney filed May 8, 2001).

Table 2

Table 2 has been amended to clarify that SEQ ID NO:12 shows homology to sequence AAR32188 of the GENESEQ database. In the application as originally filed, the sequence homologous to SEQ ID NO:12 was listed in Table 2 as R32188. This amendment clarifies which database and sequence entry were intended. No new matter is added to the specification by this amendment. Therefore, entry of this amendment is respectfully requested.

Comments regarding restriction requirement

Applicants affirm the election with traverse of claims 21-31, and 35-37, corresponding to the invention of Group A, and the election with traverse of SEQ ID NO:12 and SEQ ID NO:25. Claims 32-34, 41, 42, and 38-40 are "method of use" claims which depend from the elected polynucleotide and polypeptide product claims. Therefore, upon allowance of the polynucleotide and polypeptide product claims, it is believed that claims 32-34, 41, 42, and 38-40 should be rejoined and considered in accordance with the Commissioner's Notice in the Official Gazette of March 26, entitled "Guidance on Treatment of Product and Process Claims in light of *In re Ochiai*, *In re Brouwer* and 35 U.S.C. § 103(b)."

Objections to the claims

Claims 21, 22, 25, 29, 30, and 36 have been amended to address the objections. These claims, as currently pending, refer only to the elected polynucleotides and polypeptides of Group A. Withdrawal of the objection is therefore requested.

Utility Rejections under 35 U.S.C. §101 and §112, First Paragraph

Claims 21-31 and 35-37 are rejected under 35 U.S.C. §§ 101 and 112, first paragraph, based on the allegation that the claimed invention lacks patentable utility. The Office Action alleges in particular that "the claimed invention is not supported by either a specific and substantial utility or a well established utility" (Office Action, page 5). Applicants respectfully traverse the rejections.

The rejection of claims 21-31 and 35-37 is improper, as the inventions of those claims have a patentable utility as set forth in the instant specification, and/or a utility well known to one of ordinary skill in the art.

The invention at issue comprises polynucleotides expressed in hematopoietic/immune, gastrointestinal, cardiovascular, and reproductive tissues (Specification, e.g., at page 6 and Table 3). The invention also comprises polypeptides encoded by the claimed polynucleotides. The claimed polypeptides are identified in the patent application as human cell surface receptor proteins, abbreviated as HCSRPs. As such, the claimed invention has numerous practical, beneficial uses in toxicology testing,

drug development, and the diagnosis of disease, none of which require knowledge of how the polypeptide actually functions.

The similarity of the claimed polypeptide to another polypeptide of known, undisputed utility by itself demonstrates utility beyond the reasonable probability required by law. HCSR-12 is, in that regard, homologous to non-CD4 glycoprotein gp120 receptor (GENESEQ AAR32188) (Specification, e.g., at Table 2). In particular, SEQ ID NO:12 shares 86% sequence identity with the gp120 receptor.

This is more than enough homology to demonstrate a reasonable probability that the utility of the gp120 receptor can be imputed to the claimed invention. It is well-known that the probability that two unrelated polypeptides share more than 40% sequence homology over 70 amino acid residues is exceedingly small. Brenner *et al.*, Proc. Natl. Acad. Sci. U.S.A. 95:6073-78 (1998). Given homology in excess of 40% over many more than 70 amino acid residues, the probability that the claimed polypeptide is related to the gp120 receptor is, accordingly, very high.

The fact that the claimed polypeptide is a member of the C-type lectin receptor family alone demonstrates utility. Each of the members of this class, regardless of their particular functions, are useful. There is no evidence that any member of this class of polypeptides, let alone a substantial number of them, would not have some patentable utility. It follows that there is a more than substantial likelihood that the claimed polypeptide also has patentable utility, regardless of its actual function. The law has never required a patentee to prove more.

There is, in addition, direct proof of the utility of the claimed invention. Applicants submit with this paper the Declarations of Bedilion and Furness describing some of the practical uses of the claimed invention in gene and protein expression monitoring applications as they would have been understood at the time of the patent application.

The Bedilion Declaration describes, in particular, how the claimed expressed polynucleotide can be used in gene expression monitoring applications that were well-known at the time the patent application was filed, and how those applications are useful in developing drugs and monitoring their activity. Dr. Bedilion states that the claimed invention is a useful tool when employed as a highly specific probe in a cDNA microarray:

Persons skilled in the art would appreciate that cDNA microarrays that contained the SEQ ID NO:12-encoding polynucleotides would be a more useful tool than cDNA microarrays that did not contain the polynucleotides in connection with conducting gene expression monitoring studies on proposed (or actual) drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders for such purposes as evaluating their efficacy and toxicity.

The Patent Examiner does not dispute that the claimed polynucleotide can be used as a probe in cDNA microarrays and used in gene expression monitoring applications. Instead, the Patent Examiner contends that the claimed polynucleotide cannot be useful without precise knowledge of its biological function. But the law never has required knowledge of biological function to prove utility. It is the claimed invention's uses, not its functions, that are the subject of a proper analysis under the utility requirement.

In any event, as demonstrated by the Bedilion Declaration, the person of ordinary skill in the art can achieve beneficial results from the claimed polynucleotide in the absence of any knowledge as to the precise function of the protein encoded by it. The uses of the claimed polynucleotide in gene expression monitoring applications are in fact independent of its precise function.

The Furness Declaration describes, in particular, how the claimed polypeptide can be used in protein expression analysis techniques such as 2-D PAGE gels and western blots. Using the claimed invention with these techniques, persons of ordinary skill in the art can better assess, for example, the potential toxic affect of a drug candidate. (Furness Declaration at ¶ [11]).

The Patent Examiner does not dispute that the claimed polypeptide can be used in 2-D PAGE gels and western blots to perform drug toxicity testing. Instead, the Patent Examiner contends that the claimed polypeptide cannot be useful without precise knowledge of its function. But the law never has required knowledge of biological function to prove utility. It is the claimed invention's uses, not its functions, that are the subject of a proper analysis under the utility requirement.

In any event, as demonstrated by the Furness Declaration, the person of ordinary skill in the art can achieve beneficial results from the claimed polypeptide in the absence of any knowledge as to the precise function of the protein. The uses of the claimed polypeptide for gene expression monitoring applications including toxicology testing are in fact independent of its precise function.

I. The Applicable Legal Standard

To meet the utility requirement of sections 101 and 112 of the Patent Act, the patent applicant need only show that the claimed invention is "practically useful," *Anderson v. Natta*, 480 F.2d 1392, 1397, 178 USPQ 458 (CCPA 1973) and confers a "specific benefit" on the public. *Brenner v. Manson*, 383 U.S. 519, 534-35, 148 USPQ 689 (1966). As discussed in a recent Court of Appeals for the Federal Circuit case, this threshold is not high:

An invention is "useful" under section 101 if it is capable of providing some identifiable benefit. See *Brenner v. Manson*, 383 U.S. 519, 534 [148 USPQ 689] (1966); *Brooktree Corp. v. Advanced Micro Devices, Inc.*, 977 F.2d 1555, 1571 [24 USPQ2d 1401] (Fed. Cir. 1992) ("to violate Section 101 the claimed device must be totally incapable of achieving a useful result"); *Fuller v. Berger*, 120 F. 274, 275 (7th Cir. 1903) (test for utility is whether invention "is incapable of serving any beneficial end").

Juicy Whip Inc. v. Orange Bang Inc., 51 USPQ2d 1700 (Fed. Cir. 1999).

While an asserted utility must be described with specificity, the patent applicant need not demonstrate utility to a certainty. In *Stiftung v. Renishaw PLC*, 945 F.2d 1173, 1180, 20 USPQ2d 1094 (Fed. Cir. 1991), the United States Court of Appeals for the Federal Circuit explained:

An invention need not be the best or only way to accomplish a certain result, and it need only be useful to some extent and in certain applications: "[T]he fact that an invention has only limited utility and is only operable in certain applications is not grounds for finding lack of utility." *Envirotech Corp. v. Al George, Inc.*, 730 F.2d 753, 762, 221 USPQ 473, 480 (Fed. Cir. 1984).

The specificity requirement is not, therefore, an onerous one. If the asserted utility is described so that a person of ordinary skill in the art would understand how to use the claimed invention, it is sufficiently specific. See *Standard Oil Co. v. Montedison, S.p.a.*, 212 U.S.P.Q. 327, 343 (3d Cir. 1981). The specificity requirement is met unless the asserted utility amounts to a "nebulous expression" such as "biological activity" or "biological properties" that does not convey meaningful information about the utility of what is being claimed. *Cross v. Iizuka*, 753 F.2d 1040, 1048 (Fed. Cir. 1985).

In addition to conferring a specific benefit on the public, the benefit must also be "substantial." *Brenner*, 383 U.S. at 534. A "substantial" utility is a practical, "real-world" utility. *Nelson v. Bowler*, 626 F.2d 853, 856, 206 USPQ 881 (CCPA 1980).

If persons of ordinary skill in the art would understand that there is a "well-established" utility for the claimed invention, the threshold is met automatically and the applicant need not make any showing to demonstrate utility. Manual of Patent Examination Procedure at § 706.03(a). Only if there is no "well-established" utility for the claimed invention must the applicant demonstrate the practical benefits of the invention. *Id.*

Once the patent applicant identifies a specific utility, the claimed invention is presumed to possess it. *In re Cortright*, 165 F.3d 1353, 1357, 49 USPQ2d 1464 (Fed. Cir. 1999); *In re Brana*, 51 F.3d 1560, 1566; 34 USPQ2d 1436 (Fed. Cir. 1995). In that case, the Patent Office bears the burden of demonstrating that a person of ordinary skill in the art would reasonably doubt that the asserted utility could be achieved by the claimed invention. *Id.* To do so, the Patent Office must provide evidence or sound scientific reasoning. *See In re Langer*, 503 F.2d 1380, 1391-92, 183 USPQ 288 (CCPA 1974). If and only if the Patent Office makes such a showing, the burden shifts to the applicant to provide rebuttal evidence that would convince the person of ordinary skill that there is sufficient proof of utility. *Brana*, 51 F.3d at 1566. The applicant need only prove a "substantial likelihood" of utility; certainty is not required. *Brenner*, 383 U.S. at 532.

II. Uses of the claimed polypeptides and polynucleotides for diagnosis of conditions and disorders characterized by expression of HCSR, for toxicology testing, and for drug discovery are sufficient utilities under 35 U.S.C. §§ 101 and 112, first paragraph

The claimed invention meets all of the necessary requirements for establishing a credible utility under the Patent Law: There are "well-established" uses for the claimed invention known to persons of ordinary skill in the art, and there are specific practical and beneficial uses for the invention disclosed in the patent application's specification. These uses are explained, in detail, in the Bedilion Declaration and the Furness Declaration accompanying this paper. Objective evidence, not considered by the Patent Office, further corroborates the credibility of the asserted utilities.

A. The use of HCSR for toxicology testing, drug discovery, and disease diagnosis are practical uses that confer "specific benefits" to the public

The claimed invention has specific, substantial, real-world utility by virtue of its use in toxicology testing, drug development and disease diagnosis through gene expression profiling. These uses are explained in detail in the accompanying Bedilion Declaration and Furness Declaration, the substance of which is not rebutted by the Patent Examiner. There is no dispute that the claimed polynucleotide is in fact a useful tool in cDNA microarrays used to perform gene expression analysis and that the claimed polypeptide is a useful tool in two-dimensional polyacrylamide gel electrophoresis ("2-D PAGE") analysis and western blots used to monitor protein expression and assess drug toxicity. These uses are sufficient to establish utilities for the claimed polynucleotide and polypeptide, respectively.

The instant application is a divisional of, and claims priority to, United States Provisional Patent Application Serial No. 60/123,404 filed on March 8, 1999 (hereinafter "the Tang et al. '404 application").

1. The Bedilion Declaration

In his Declaration, Dr. Bedilion explains the many reasons why a person skilled in the art reading the Tang et al. '404 application on March 8, 1999 would have understood that application to disclose the claimed polynucleotide to be useful for a number of gene expression monitoring applications, *e.g.*, as a highly specific probe for the expression of that specific polynucleotide in connection with the development of drugs and the monitoring of the activity of such drugs. (Bedilion Declaration at, *e.g.*, ¶¶ 10-15). Much, but not all, of Dr. Bedilion's explanation concerns the use of the claimed polynucleotide in cDNA microarrays of the type first developed at Stanford University for evaluating the efficacy and toxicity of drugs, as well as for other applications. (Bedilion Declaration, ¶¶ 12 and 15).¹

In connection with his explanations, Dr. Bedilion states that the "Tang et al. '404 specification

¹Dr. Bedilion also explained, for example, why persons skilled in the art would also appreciate, based on the Tang et al. '404 specification, that the claimed polynucleotide would be useful in connection with developing new drugs using technology, such as Northern analysis, that predated by

would have led a person skilled in the art on March 8, 1999 who was using gene expression monitoring in connection with working on developing new drugs for the treatment of cell proliferative disorders, immune system disorders, infections, and neuronal disorders [a] to conclude that a cDNA microarray that contained the SEQ ID NO:12-encoding polynucleotides would be a highly useful tool, and [b] to request specifically that any cDNA microarray that was being used for such purposes contain the SEQ ID NO:12-encoding polynucleotides" (Bedilion Declaration, ¶ 15). For example, as explained by Dr. Bedilion, "[p]ersons skilled in the art would [have appreciated on March 8, 1999] that a cDNA microarray that contained the SEQ ID NO:12-encoding polynucleotides would be a more useful tool than a cDNA microarray that did not contain the polynucleotides in connection with conducting gene expression monitoring studies on proposed (or actual) drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders for such purposes as evaluating their efficacy and toxicity." *Id.*

In support of those statements, Dr. Bedilion provided detailed explanations of how cDNA technology can be used to conduct gene expression monitoring evaluations, with extensive citations to pre-March 8, 1999 publications showing the state of the art on March 8, 1999. (Bedilion Declaration, ¶¶ 10-14). While Dr. Bedilion's explanations in paragraph 15 of his Declaration include more than three pages of text and six subparts (a)-(f), he specifically states that his explanations are not "all-inclusive." *Id.* For example, with respect to toxicity evaluations, Dr. Bedilion had earlier explained how persons skilled in the art who were working on drug development on March 8, 1999 (and for several years prior to March 8, 1999) "without any doubt" appreciated that the toxicity (or lack of toxicity) of any proposed drug was "one of the most important criteria to be evaluated in connection with the development of the drug" and how the teachings of the Tang et al. '404 application clearly include using differential gene expression analyses in toxicity studies (Bedilion Declaration, ¶ 10).

Thus, the Bedilion Declaration establishes that persons skilled in the art reading the Tang et al. '404 application at the time it was filed "would have wanted their cDNA microarray to have a [SEQ ID NO:12-encoding polynucleotide probe] because a microarray that contained such a probe (as compared to one that did not) would provide more useful results in the kind of gene expression monitoring studies using cDNA microarrays that persons skilled in the art have been doing since well

prior to March 8, 1999" (Bedilion Declaration, ¶ 15, item (f)). This, by itself, provides more than sufficient reason to compel the conclusion that the Tang et al. '404 application disclosed to persons skilled in the art at the time of its filing substantial, specific and credible real-world utilities for the claimed polynucleotide.

Nowhere does the Patent Examiner address the fact that, as described on p. 33 of the Tang et al. '404 application, the claimed polynucleotides can be used as highly specific probes in, for example, cDNA microarrays – probes that without question can be used to measure both the existence and amount of complementary RNA sequences known to be the expression products of the claimed polynucleotides. The claimed invention is not, in that regard, some random sequence whose value as a probe is speculative or would require further research to determine.

Given the fact that the claimed polynucleotide is known to be expressed, its utility as a measuring and analyzing instrument for expression levels is as indisputable as a scale's utility for measuring weight. This use as a measuring tool, regardless of how the expression level data ultimately would be used by a person of ordinary skill in the art, by itself demonstrates that the claimed invention provides an identifiable, real-world benefit that meets the utility requirement. *Raytheon v. Roper*, 724 F.2d 951, (Fed. Cir. 1983) (claimed invention need only meet one of its stated objectives to be useful); *In re Cortwright*, 165 F.3d 1353, 1359 (Fed. Cir. 1999) (how the invention works is irrelevant to utility); MPEP § 2107 ("Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific, and unquestionable utility (e.g., they are useful in analyzing compounds)" (emphasis added)).

The Bedilion Declaration shows that a number of pre-March 8, 1999 publications confirm and further establish the utility of cDNA microarrays in a wide range of drug development gene expression monitoring applications at the time the Tang et al. '404 application was filed (Bedilion Declaration ¶¶ 10-14; Bedilion Exhibits A-G). Indeed, Brown and Shalon U.S. Patent No. 5,807,522 (the Brown '522 patent, Bedilion Exhibit D), which issued from a patent application filed in June 1995 and was effectively published on December 29, 1995 as a result of the publication of a PCT counterpart application, shows that the Patent Office recognizes the patentable utility of the cDNA technology

developed in the early to mid-1990s. As explained by Dr. Bedilion, among other things (Bedilion Declaration, ¶ 12):

The Brown '522 patent further teaches that the "[m]icroarrays of immobilized nucleic acid sequences prepared in accordance with the invention" can be used in "numerous" genetic applications, including "monitoring of gene expression" applications (see Bedilion Tab D at col. 14, lines 36-42). The Brown '522 patent teaches (a) monitoring gene expression (i) in different tissue types, (ii) in different disease states, and (iii) in response to different drugs, and (b) that arrays disclosed therein may be used in toxicology studies (see Bedilion Tab D at col. 15, lines 13-18 and 52-58 and col. 18, lines 25-30).

Literature reviews published shortly after the filing of the Tang et al. '404 application describing the state of the art further confirm the claimed invention's utility. Rockett et al. confirm, for example, that the claimed invention is useful for differential expression analysis regardless of how expression is regulated:

Despite the development of multiple technological advances which have recently brought the field of gene expression profiling to the forefront of molecular analysis, recognition of the importance of differential gene expression and characterization of differentially expressed genes has existed for many years.

* * *

Although differential expression technologies are applicable to a broad range of models, perhaps their most important advantage is that, in most cases, absolutely no prior knowledge of the specific genes which are up- or down-regulated is required.

* * *

Whereas it would be informative to know the identity and functionality of all genes up/down regulated by . . . toxicants, this would appear a longer term goal However, the current use of gene profiling yields a *pattern* of gene changes for a xenobiotic of unknown toxicity which may be matched to that of well characterized toxins, thus alerting the toxicologist to possible *in vivo* similarities between the unknown and the standard, thereby providing a platform for more extensive toxicological examination. (emphasis added)

Rockett et al., Differential gene expression in drug metabolism and toxicology: practicalities, problems

toxicology 20 (N.Y. Acad. Sci. Ser. 7) 655 (1999)

In another pre-March 8, 1999 article, Lashkari et al. state explicitly that sequences that are merely "predicted" to be expressed (predicted Open Reading Frames, or ORFs) – the claimed invention in fact is known to be expressed – have numerous uses:

Efforts have been directed toward the amplification of each predicted ORF or any other region of the genome ranging from a few base pairs to several kilobase pairs. There are many uses for these amplicons– they can be cloned into standard vectors or specialized expression vectors, or can be cloned into other specialized vectors such as those used for two-hybrid analysis. The amplicons can also be used directly by, for example, arraying onto glass for expression analysis, for DNA binding assays, or for any direct DNA assay.

Lashkari et al., Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR, 94 Proc. Nat. Acad. Sci. 8945 (Aug. 1997) (emphasis added).

2. The Furness Declaration

In his Declaration, Mr. Furness explains the many reasons why a person skilled in the art who read the Tang et al. '404 application on March 8, 1999 would have understood that application to disclose the claimed polypeptide to be useful for a number of gene and protein expression monitoring applications, *e.g.*, in 2-D PAGE technologies, in connection with the development of drugs and the monitoring of the activity of such drugs. (Furness Declaration at, *e.g.*, ¶¶ [11-15]). Much, but not all, of Mr. Furness' explanation concerns the use of the claimed polypeptide in the creation of protein expression maps using 2-D PAGE.

2-D PAGE technologies were developed during the 1980's. Since the early 1990's, 2-D PAGE has been used to create maps showing the differential expression of proteins in different cell types or in similar cell types in response to drugs and potential toxic agents. Each expression pattern reveals the state of a tissue or cell type in its given environment, *e.g.*, in the presence or absence of a drug. By comparing a map of cells treated with a potential drug candidate to a map of cells not treated with the candidate, for example, the potential toxicity of a drug can be assessed. Furness Declaration at ¶ [11].)

The claimed invention makes 2-D PAGE analysis a more powerful tool for toxicology and drug efficacy testing. A person of ordinary skill in the art can derive more information about the state or states or tissue or cell samples from 2-D PAGE analysis with the claimed invention than without it. As Mr. Furness explains:

In view of the Tang et al. '404 application, the Wilkins article, and other related pre-March, 1999 publications, persons skilled in the art on March 8, 1999 clearly would have understood the Tang et al. '404 application to disclose the SEQ ID NO:12 polypeptide to be useful in 2-D PAGE analyses for the development of new drugs and monitoring the activities of drugs for such purposes as evaluating their efficacy and toxicity. . . . (Furness Declaration, ¶10)

* * *

Persons skilled in the art would appreciate that a 2-D PAGE map that utilized the SEQ ID NO:12 polypeptide sequence would be a more useful tool than a 2-D PAGE map that did not utilize this protein sequence in connection with conducting protein expression monitoring studies on proposed (or actual) drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders for such purposes as evaluating their efficacy and toxicity. (Furness Declaration, ¶12)

Mr. Furness' observations are confirmed in the literature published before the filing of the patent application. Wilkins, for example, describes how 2-D gels are used to define proteins present in various tissues and measure their levels of expression, the data from which is in turn used in databases:

For proteome projects, the aim of [computer-aided 2-D PAGE] analysis . . . is to catalogue all spots from the 2-D gel in a qualitative and if possible quantitative manner, so as to define the number of proteins present and their levels of expression. Reference gel images, constructed from one or more gels, for the basis of two-dimensional gel databases. (Wilkins, Tab C, p. 26).

B. The use of nucleic acids coding for proteins expressed by humans as tools for toxicology testing, drug discovery, and the diagnosis of disease is now "well-established"

The technologies made possible by expression profiling and the DNA tools upon which they rely are now well-established. The technical literature recognizes not only the prevalence of these

technologies, but also their unprecedented advantages in drug development, testing and safety assessment. These technologies include toxicology testing, as described by Bedilion in his Declaration.

Toxicology testing is now standard practice in the pharmaceutical industry. See, *e.g.*, John C. Rockett et al., *supra*:

Knowledge of toxin-dependent regulation in target tissues is not solely an academic pursuit as much interest has been generated in the pharmaceutical industry to harness this technology in the early identification of toxic drug candidates, thereby shortening the developmental process and contributing substantially to the safety assessment of new drugs.

To the same effect are several other scientific publications, including Emile F. Nuwaysir et al.,

Microarrays and Toxicology: The Advent of Toxicogenomics, 24 Molecular Carcinogenesis 153 (1999); Sandra Steiner and N. Leigh Anderson, Expression profiling in toxicology -- potentials and limitations, 112-13 Toxicology Letters 467 (2000).

Nucleic acids useful for measuring the expression of whole classes of genes are routinely incorporated for use in toxicology testing. Nuwaysir et al. describes, for example, a Human ToxChip comprising 2089 human clones, which were selected

for their well-documented involvement in basic cellular processes as well as their responses to different types of toxic insult. Included on this list are DNA replication and repair genes, apoptosis genes, and genes responsive to PAHs and dioxin-like compounds, peroxisome proliferators, estrogenic compounds, and oxidant stress. Some of the other categories of genes include transcription factors, oncogenes, tumor suppressor genes, cyclins, kinases, phosphatases, cell adhesion and motility genes, and homeobox genes. Also included in this group are 84 housekeeping genes, whose hybridization intensity is averaged and used for signal normalization of the other genes on the chip.

See also Table 1 of Nuwaysir et al. (listing additional classes of genes deemed to be of special interest in making a human toxicology microarray).

The more genes that are available for use in toxicology testing, the more powerful the technique. "Arrays are at their most powerful when they contain the entire genome of the species they are being used to study." John C. Rockett and David J. Dix, Application of DNA Arrays to Toxicology, 107 Environ. Health Perspec. 681, No. 8 (1999). Control genes are carefully selected for their stability across a large set of array experiments in order to best study the effect of toxicological compounds. See attached email from the primary investigator on the Nuwaysir paper, Dr. Cynthia Afshari, to an

Incyte employee, dated July 3, 2000, as well as the original message to which she was responding, indicating that even the expression of carefully selected control genes can be altered. Thus, there is no expressed gene which is irrelevant to screening for toxicological effects, and all expressed genes have a utility for toxicological screening.

In fact, the potential benefit to the public, in terms of lives saved and reduced health care costs, are enormous. Recent developments provide evidence that the benefits of this information are already beginning to manifest themselves. Examples include the following:

- In 1999, CV Therapeutics, an Incyte collaborator, was able to use Incyte gene expression technology, information about the structure of a known transporter gene, and chromosomal mapping location, to identify the key gene associated with Tangiers disease. This discovery took place over a matter of only a few weeks, due to the power of these new genomics technologies. The discovery received an award from the American Heart Association as one of the top 10 discoveries associated with heart disease research in 1999.
- In an April 9, 2000, article published by the Bloomberg news service, an Incyte customer stated that it had reduced the time associated with target discovery and validation from 36 months to 18 months, through use of Incyte's genomic information database. Other Incyte customers have privately reported similar experiences. The implications of this significant saving of time and expense for the number of drugs that may be developed and their cost are obvious.
- In a February 10, 2000, article in the *Wall Street Journal*, one Incyte customer stated that over 50 percent of the drug targets in its current pipeline were derived from the Incyte database. Other Incyte customers have privately reported similar experiences. By doubling the number of targets available to pharmaceutical researchers, Incyte genomic information has demonstrably accelerated the development of new drugs.

C. The use of proteins expressed by humans as tools for toxicology testing, drug discovery, and the diagnosis of disease is now "well-established"

The technologies made possible by expression profiling using polypeptides are now well-established. The technical literature recognizes not only the prevalence of these technologies, but also their unprecedented advantages in drug development, testing and safety assessment. These technologies include toxicology testing, as described by Furness in his Declaration.

Toxicology testing is now standard practice in the pharmaceutical industry. See, *e.g.*, John C. Rockett, *et al.*, Differential gene expression in drug metabolism and toxicology: practicalities, problems, and potential, *Xenobiotica* 29:655-691 (July 1999) (Reference No. 1):

Knowledge of toxin-dependent regulation in target tissues is not solely an academic pursuit as much interest has been generated in the pharmaceutical industry to harness this technology in the early identification of toxic drug candidates, thereby shortening the developmental process and contributing substantially to the safety assessment of new drugs. ((Reference No. 1), page 656)

To the same effect are several other scientific publications, including Emile F. Nuwaysir, *et al.*, Microarrays and Toxicology: The Advent of Toxicogenomics, *Molecular Carcinogenesis* 24:153-159 (1999) (Reference No. 2); Sandra Steiner and N. Leigh Anderson, Expression profiling in toxicology - potentials and limitations, *Toxicology Letters* 112-13:467-471 (2000) (Reference No. 3).

The more genes – and, accordingly, the polypeptides they encode -- that are available for use in toxicology testing, the more powerful the technique. Control genes are carefully selected for their stability across a large set of array experiments in order to best study the effect of toxicological compounds. See attached email from the primary investigator, Dr. Cynthia Afshari to an Incyte employee, dated July 3, 2000, as well as the original message to which she was responding (Reference No. 4) Thus, there is no expressed gene which is irrelevant to screening for toxicological effects, and all expressed genes have a utility for toxicological screening.

In fact, the potential benefit to the public, in terms of lives saved and reduced health care costs, are enormous. Recent developments provide evidence that the benefits of this information are already beginning to manifest themselves. Examples include the following:

- In 1999, CV Therapeutics, an Incyte collaborator, was able to use Incyte gene expression technology, information about the structure of a known transporter gene, and chromosomal mapping location, to identify the key gene associated with Tangier disease. This discovery took place over a matter of only a few weeks, due to the power of these new genomics technologies. The discovery received an award from the American Heart Association as one of the top 10 discoveries associated with heart disease research in 1999.
- In an April 9, 2000, article published by the Bloomberg news service, an Incyte customer stated that it had reduced the time associated with target discovery and validation from 36 months to 18 months, through use of Incyte's genomic information

database. Other Incyte customers have privately reported similar experiences. The implications of this significant saving of time and expense for the number of drugs that may be developed and their cost are obvious.

- In a February 10, 2000, article in the *Wall Street Journal*, one Incyte customer stated that over 50 percent of the drug targets in its current pipeline were derived from the Incyte database. Other Incyte customers have privately reported similar experiences. By doubling the number of targets available to pharmaceutical researchers, Incyte genomic information has demonstrably accelerated the development of new drugs.

Because the Patent Examiner failed to address or consider the "well-established" utilities for the claimed invention in toxicology testing, drug development, and the diagnosis of disease, the Examiner's rejections should be overturned regardless of their merit.

D. The similarity of the claimed polypeptide to another of undisputed utility demonstrates utility

Because there is a substantial likelihood that the claimed HCSRП is functionally related to the gp120 receptor, a polypeptide of undisputed utility, there is by implication a substantial likelihood that the claimed polypeptide and the polynucleotide that encodes it are similarly useful. Applicants need not show any more to demonstrate utility. *In re Brana*, 51 F.3d at 1567.

It is undisputed that the claimed polypeptide shown as SEQ ID NO:12 in the patent application and referred to as HCSRП-12 shares more than 85% sequence identity over 325 amino acid residues with the gp120 receptor (GENESEQ AAR32188, International Patent WO 93/01820). The gp120 receptor belongs to the C-lectin receptor family whose members are known to mediate cellular immunity in part through carbohydrate recognition on microorganisms. Members of this family have been shown to bind glycoproteins on the viral envelopes of human immunodeficiency virus (HIV) and Ebola virus (enclosed references of Curtis et al., Turville et al., and Alvarez et al.). Indeed, HCSRП-12 shows homology to other members of the C-lectin receptor family. The Examiner's attention is directed to Exhibit B, which shows a recent BLAST analysis of SEQ ID NO:12. The top hits include human L-SIGN (g13383470 and human mDC-SIGN type I isoform (g1538306), which except for a

few sequence insertions, share 99.7% identity with SEQ ID NO:12. Both L-SIGN and DC-SIGN are known to bind to HIV gp120 and Ebola virus glycoproteins (enclosed references of Turville et al., Bashirova et al., and Alvarez et al.). The alignment of HCSR-12 with human L-SIGN and mDC-SIGN corroborates the original determination of the instant application that HCSR-12 was correctly assigned to the class of receptors that bind to HIV envelope glycoprotein gp120.

The attention of the Examiner is directed to Exhibit C which shows that SEQ ID NO:12 contains a C-type lectin domain from about residue S211 to residue K317 as determined by recent HMMER, MOTIFS, and BLIMPS analyses. Exhibit D shows an alignment of SEQ ID NO:12 with the sequences of the gp120 receptor (GENESEQ AAR32188), a membrane-associated C-type lectin that binds human immunodeficiency virus envelope glycoprotein gp120 (g8572543), and L-SIGN (g13383470), performed using the program, MEGALIGN version 4.05 and the CLUSTAL V algorithm. This alignment shows the presence of conserved residues, particularly in the region of SEQ ID NO:12 corresponding to the lectin domain. In all of these proteins, a C-type lectin domain is believed to mediate carbohydrate recognition and binding to the HIV envelope glycoprotein, gp120.

The homology among these sequences is more than enough to demonstrate a reasonable probability that the utility of the gp120 receptor can be imputed to the claimed invention. It is well-known that the probability that two unrelated polypeptides share more than 40% sequence homology over 70 amino acid residues is exceedingly small (Brenner et. al., Proc. Natl. Acad. Sci. (1998) 95:6073-78). Given homology in excess of 40% over many more than 70 amino acid residues, the probability that the claimed polypeptide is related to the gp120 receptor is, accordingly, very high.

It was known in the art at the time the application was filed that C-lectin receptors such as the gp120 receptor could be useful for detection of virus, inhibition of viral infection, and for development of vaccines (enclosed references of Geijtenbeek et al. and International patent WO 93/01820). It was also known that infection with HIV is associated with an increased incidence of cancer, particularly with Kaposi's sarcoma and non-Hodgkin's lymphoma, and that gp120 plays a role in tumor metastasis (enclosed references of Scadden and Hodgson et al.). In addition, gp120 induces neuronal apoptosis and neuronal injury associated with neurodegenerative disorders caused by HIV infection (enclosed references of Kaul et al. and

the gp120 receptor and C-lectin receptor proteins as a class, persons skilled in the art at the time the application was filed would have considered HCSR-12 to be an important and valuable tool for use in research on cell proliferative disorders, immune system disorders, infections, and neuronal disorders.

The Examiner must accept the Applicants' demonstration that the homology between the claimed invention and the gp120 receptor demonstrates utility by a reasonable probability unless the Examiner can demonstrate through evidence or sound scientific reasoning that a person of ordinary skill in the art would doubt utility. *See In re Langer*, 503 F.2d 1380, 1391-92, 183 USPQ 288 (CCPA 1974). The Examiner has not provided sufficient evidence or sound scientific reasoning to the contrary.

E. Objective evidence corroborates the utilities of the claimed invention

There is, in fact, no restriction on the kinds of evidence a Patent Examiner may consider in determining whether a "real-world" utility exists. Indeed, "real-world" evidence, such as evidence showing actual use or commercial success of the invention, can demonstrate conclusive proof of utility. *Raytheon v. Roper*, 220 USPQ2d 592 (Fed. Cir. 1983); *Nestle v. Eugene*, 55 F.2d 854, 856, 12 USPQ 335 (6th Cir. 1932). Indeed, proof that the invention is made, used or sold by any person or entity other than the patentee is conclusive proof of utility. *United States Steel Corp. v. Phillips Petroleum Co.*, 865 F.2d 1247, 1252, 9 USPQ2d 1461 (Fed. Cir. 1989).

Over the past several years, a vibrant market has developed for databases containing all expressed genes (along with the polypeptide translations of those genes), in particular genes having medical and pharmaceutical significance such as the instant sequence. (Note that while the value in these databases is enhanced by their completeness, each sequence in them is independently valuable nonetheless.) The databases sold by Applicants' assignee, Incyte, include exactly the kinds of information made possible by the claimed invention, such as tissue and disease associations. Incyte sells its database containing the claimed sequence and millions of other sequences throughout the scientific community, including to pharmaceutical companies who use the information to develop new pharmaceuticals.

Both Incyte's customers and the scientific community have acknowledged that Incyte's databases have proven to be valuable in, for example, the identification and development of drug candidates. As Incyte adds information to its databases, including the information that can be generated only as a result of Incyte's discovery of the claimed polynucleotide and its use of that polynucleotide on cDNA microarrays, the databases become even more powerful tools. Thus the claimed invention adds more than incremental benefit to the drug discovery and development process.

III. The Patent Examiner's Rejections Are Without Merit

Rather than responding to the evidence demonstrating utility, the Examiner attempts to dismiss it altogether by arguing that the disclosed and well-established utilities for the claimed polynucleotide and polypeptide are not "specific and substantial" utilities. (Office Action at p. 5.) The Examiner is incorrect both as a matter of law and as a matter of fact.

A. The Precise Biological Role Or Function Of An Expressed Polynucleotide or Polypeptide Is Not Required To Demonstrate Utility

The Patent Examiner's primary rejection of the claimed invention is based on the ground that, without information as to the precise "biological role" of the claimed invention, the claimed invention's utility is not sufficiently specific. According to the Examiner, it is not enough that a person of ordinary skill in the art could use and, in fact, would want to use the claimed invention either by itself or in a microarray, 2-D gel or western blot to monitor the expression of genes for such applications as the evaluation of a drug's efficacy and toxicity. The Examiner would require, in addition, that the applicant provide a specific and substantial interpretation of the results generated in any given expression analysis.

It may be that specific and substantial interpretations and detailed information on biological function are necessary to satisfy the requirements for publication in some technical journals, but they are not necessary to satisfy the requirements for obtaining a United States patent. The relevant question is not, as the Examiner would have it, whether it is known how or why the invention works, *In re Cortwright*, 165 F.3d 1353, 1359 (Fed. Cir. 1999), but rather whether the invention provides an

1364, 1366 (Fed. Cir. 1999). If the benefit exists, and there is a substantial likelihood the invention provides the benefit, it is useful. There can be no doubt, particularly in view of the Bedilion Declaration (at, *e.g.*, ¶¶ 10 and 15, Bedilion) and the Furness Declaration (at, *e.g.*, ¶¶ 10-13), that the present invention meets this test.

The threshold for determining whether an invention produces an identifiable benefit is low. *Juicy Whip*, 185 F.3d at 1366. Only those utilities that are so nebulous that a person of ordinary skill in the art would not know how to achieve an identifiable benefit and, at least according to the PTO guidelines, so-called "throwaway" utilities that are not directed to a person of ordinary skill in the art at all, do not meet the statutory requirement of utility. Utility Examination Guidelines, 66 Fed. Reg. 1092 (Jan. 5, 2001).

Knowledge of the biological function or role of a biological molecule has never been required to show real-world benefit. In its most recent explanation of its own utility guidelines, the PTO acknowledged so much (66 F.R. at 1095):

[T]he utility of a claimed DNA does not necessarily depend on the function of the encoded gene product. A claimed DNA may have specific and substantial utility because, *e.g.*, it hybridizes near a disease-associated gene or it has gene-regulating activity.

By implicitly requiring knowledge of biological function for any claimed nucleic acid, the Examiner has, contrary to law, elevated what is at most an evidentiary factor into an absolute requirement of utility. Rather than looking to the biological role or function of the claimed invention, the Examiner should have looked first to the benefits it is alleged to provide.

B. Membership in a Class of Useful Products Can Be Proof of Utility

Despite the uncontradicted evidence that the claimed polypeptide is related to the gp120 receptor, a member of the C-type lectin receptor family, whose members indisputably are useful, the Examiner refused to impute the utility of the gp120 receptor to HCSR-12. In the Office Action of January 15, 2003, the Patent Examiner takes the position that unless Applicants can identify which

particular biological function of the gp120 receptor is possessed by HCSR-12, utility cannot be imputed.

In order to demonstrate utility by membership in a class, the law requires only that the class not contain a substantial number of useless members. So long as the class does not contain a substantial number of useless members, there is sufficient likelihood that the claimed invention will have utility, and a rejection under 35 U.S.C. § 101 is improper. That is true regardless of how the claimed invention ultimately is used and whether or not the members of the class possess one utility or many. See *Brenner v. Manson*, 383 U.S. 519, 532 (1966); *Application of Kirk*, 376 F.2d 936, 943 (CCPA 1967).

Membership in a "general" class is insufficient to demonstrate utility only if the class contains a sufficient number of useless members such that a person of ordinary skill in the art could not impute utility by a substantial likelihood. There would be, in that case, a substantial likelihood that the claimed invention is one of the useless members of the class. In the few cases in which class membership did not prove utility by substantial likelihood, the classes did in fact include predominately useless members. E.g., *Brenner* (man-made steroids); *Kirk* (same); *Natta* (man-made polyethylene polymers).

The Examiner addresses HCSR-12 as if the general class in which it is included is not the C-type lectin receptor family, but rather all polynucleotides or all polypeptides, including the vast majority of useless theoretical molecules not occurring in nature, and thus not pre-selected by nature to be useful. While these "general classes" may contain a substantial number of useless members, the C-type lectin receptor family does not. The C-type lectin receptor family is sufficiently specific to rule out any reasonable possibility that HCSR-12 would not also be useful like the other members of the family.

Because the Examiner has not presented any evidence that the C-type lectin receptor class of proteins has any, let alone a substantial number, of useless members, the Examiner must conclude that there is a "substantial likelihood" that the HCSR-12 encoded by the claimed polynucleotides is useful. It follows that SEQ ID NO:12 and SEQ ID NO:25 also are useful.

Even if the Examiner's "common utility" criterion were correct – and it is not – the gp120 receptor would meet it. It is undisputed that known members of the C-type lectin receptor family, including the gp120 receptor, function in cellular immunity and host defense against viral infections. A

person of ordinary skill in the art need not know any more about how the claimed invention functions in cellular immunity and viral infections to use it, and the Examiner presents no evidence to the contrary. Instead, the Examiner makes the conclusory observation that a person of ordinary skill in the art would need to know whether, for example, any given gp120 receptor functions in cellular immunity or viral infections. The Examiner then goes on to assume that the only use for HCSR-12 absent knowledge as to how HCSR-12 actually works is further study of HCSR-12 itself.

Not so. As demonstrated by Applicants, knowledge that HCSR-12 is a C-type lectin related to the gp120 receptor is more than sufficient to make it useful for the diagnosis and treatment of cell proliferative disorders, immune system disorders, infections, and neuronal disorders. The Examiner must accept these facts to be true unless the Examiner can provide evidence or sound scientific reasoning to the contrary. But the Examiner has not done so.

C. Because the uses of polynucleotides encoding HCSR in toxicology testing, drug discovery, and disease diagnosis are practical uses beyond mere study of the invention itself, the claimed invention has substantial utility.

The Examiner rejected the claims at issue on the ground that the use of an invention as tool for research is not a "substantial" use. Because the Examiner's rejection assumes a substantial overstatement of the law, and is incorrect in fact, it must be overturned.

There is no authority for the proposition that use as a tool for research is not a substantial utility. Indeed, the Patent Office has recognized that just because an invention is used in a research setting does not mean that it lacks utility (MPEP § 2107):

Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific and unquestionable utility (e.g., they are useful in analyzing compounds). An assessment that focuses on whether an invention is useful only in a research setting thus does not address whether the specific invention is in fact "useful" in a patent sense. Instead, Office personnel must distinguish between inventions that have a specifically identified utility and inventions whose specific utility requires further research to identify or reasonably confirm.

The Patent Office's actual practice has been, at least until the present, consistent with that approach. It has routinely issued patents for inventions whose only use is to facilitate research, such as

DNA ligases. These are acknowledged by the PTO's Training Materials themselves to be useful, as well as DNA sequences used, for example, as markers.

Only a limited subset of research uses are not "substantial" utilities: those in which the only known use for the claimed invention is to be an **object** of further study, thus merely inviting further research. This follows from *Brenner*, in which the U.S. Supreme Court held that a process for making a compound does not confer a substantial benefit where the only known use of the compound was to be the object of further research to determine its use. *Id.* at 535. Similarly, in *Kirk*, the Court held that a compound would not confer substantial benefit on the public merely because it might be used to synthesize some other, unknown compound that would confer substantial benefit. *Kirk*, 376 F.2d at 940, 945 ("What Applicants are really saying to those in the art is take these steroids, experiment, and find what use they do have as medicines."). Nowhere do those cases state or imply, however, that a material cannot be patentable if it has some other beneficial use in research.

Such beneficial uses beyond studying the claimed invention itself have been demonstrated, in particular those described in the Bedilion and Furness Declarations. The claimed invention is a tool, rather than an object, of research. The data generated in gene expression monitoring using the claimed invention as a tool is **not** used merely to study the claimed polynucleotide itself, but rather to study properties of tissues, cells, and potential drug candidates and toxins. Without the claimed invention, the information regarding the properties of tissues, cells, drug candidates and toxins is less complete.

Moreover, as discussed above in section II D., SEQ ID NO:12 shares homology with other members of the C-lectin family that bind to viral glycoproteins. Therefore, the skilled artisan would have considered HCSRП to be an important and valuable tool, in particular, for use in research on cell proliferative disorders, immune system disorders, infections, and neuronal disorders. The claimed invention has numerous other uses as a research tool, each of which alone is a "substantial utility." These include uses such as diagnostic assays (e.g., pages 40-44), chromosomal markers (e.g., pages 44-45), ligand screening assays (e.g., page 33), and drug screening (page 45-46).

IV. By Requiring the Patent Applicant to Assert a Particular or Unique Utility, the Patent Examination Utility Guidelines and Training Materials Applied by the Patent Examiner Misstate the Law

There is an additional, independent reason to overturn the rejections: to the extent the rejections are based on Revised Interim Utility Examination Guidelines (64 FR 71427, December 21, 1999), the final Utility Examination Guidelines (66 FR 1092, January 5, 2001) and/or the Revised Interim Utility Guidelines Training Materials (USPTO Website www.uspto.gov, March 1, 2000), the Guidelines and Training Materials are themselves inconsistent with the law.

The Training Materials, which direct the Examiners regarding how to apply the Utility Guidelines, address the issue of specificity with reference to two kinds of asserted utilities: "specific" utilities which meet the statutory requirements, and "general" utilities which do not. The Training Materials define a "specific utility" as follows:

A [specific utility] is *specific* to the subject matter claimed. This contrasts to *general* utility that would be applicable to the broad class of invention. For example, a claim to a polynucleotide whose use is disclosed simply as "gene probe" or "chromosome marker" would not be considered to be specific in the absence of a disclosure of a specific DNA target. Similarly, a general statement of diagnostic utility, such as diagnosing an unspecified disease, would ordinarily be insufficient absent a disclosure of what condition can be diagnosed.

The Training Materials distinguish between "specific" and "general" utilities by assessing whether the asserted utility is sufficiently "particular," *i.e.*, unique (Training Materials at p.52) as compared to the "broad class of invention." (In this regard, the Training Materials appear to parallel the view set forth in Stephen G. Kunin, Written Description Guidelines and Utility Guidelines, 82 J.P.T.O.S. 77, 97 (Feb. 2000) ("With regard to the issue of specific utility the question to ask is whether or not a utility set forth in the specification is *particular* to the claimed invention.")).

Such "unique" or "particular" utilities never have been required by the law. To meet the utility requirement, the invention need only be "practically useful." *Natta*, 480 F.2d 1 at 1397, and confer a "specific benefit" on the public. *Brenner*, 383 U.S. at 534. Thus, incredible "throwaway" utilities, such as trying to "patent a transgenic mouse by saying it makes great snake food," do not meet this standard. Karen Hall, Genomic Warfare, *The American Lawyer* 68 (June 2000) (quoting John Doll, Chief of the Biotech Section of USPTO).

This does not preclude, however, a general utility, contrary to the statement in the Training Materials where "specific utility" is defined (page 5). Practical real-world uses are not limited to uses that are unique to an invention. The law requires that the practical utility be "definite," not particular. *Montedison*, 664 F.2d at 375. Applicant is not aware of any court that has rejected an assertion of utility on the grounds that it is not "particular" or "unique" to the specific invention. Where courts have found utility to be too "general," it has been in those cases in which the asserted utility in the patent disclosure was not a practical use that conferred a specific benefit. That is, a person of ordinary skill in the art would have been left to guess as to how to benefit at all from the invention. In *Kirk*, for example, the CCPA held the assertion that a man-made steroid had "useful biological activity" was insufficient where there was no information in the specification as to how that biological activity could be practically used. *Kirk*, 376 F.2d at 941.

The fact that an invention can have a particular use does not provide a basis for requiring a particular use. *See Brana, supra* (disclosure describing a claimed antitumor compound as being homologous to an antitumor compound having activity against a "particular" type of cancer was determined to satisfy the specificity requirement). "Particularity" is not and never has been the *sine qua non* of utility; it is, at most, one of many factors to be considered.

As described *supra*, broad classes of inventions can satisfy the utility requirement so long as a person of ordinary skill in the art would understand how to achieve a practical benefit from knowledge of the class. Only classes that encompass a significant portion of nonuseful members would fail to meet the utility requirement. *Supra* § II.B.2 (*Montedison*, 664 F.2d at 374-75).

The Training Materials fail to distinguish between broad classes that convey information of practical utility and those that do not, lumping all of them into the latter, unpatentable category of "general" utilities. As a result, the Training Materials paint with too broad a brush. Rigorously applied, they would render unpatentable whole categories of inventions that heretofore have been considered to be patentable and that have indisputably benefitted the public, including the claimed invention. *See supra* § II.B. Thus the Training Materials cannot be applied consistently with the law.

V. To the Extent the Rejection of the Patented Invention under 35 U.S.C. § 112, First Paragraph, Is Based on the Improper Rejection for Lack of Utility under 35 U.S.C. § 101, it Must Be Reversed.

The rejection set forth in the Office Action is based on the assertions discussed above, i.e., that the claimed invention lacks patentable utility. To the extent that the rejection under § 112, first paragraph, is based on the improper allegation of lack of patentable utility under § 101, it fails for the same reasons.

Written description rejections under 35 U.S.C. § 112, first paragraph

Claims 21, 23, 26, 27, 28, 30, 31, 35, and 37 have been rejected under the first paragraph of 35 U.S.C. 112 for alleged lack of an adequate written description. This rejection is respectfully traversed.

As a preliminary matter, claim 21 (b) has been amended to recite "a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:12, wherein said polypeptide binds to human immunodeficiency virus glycoprotein gp120." Claim 21 (c) has been amended to recite a biologically active fragment of a polypeptide having the amino acid sequence of SEQ ID NO:12, wherein said fragment binds to human immunodeficiency virus glycoprotein gp120. By these amendments, Applicants expressly do not disclaim equivalents of the invention which could include polypeptides and fragments having biological activities other than binding to human immunodeficiency virus glycoprotein gp120. Applicants are amending the claims solely to obtain expeditious allowance of the instant application. Support for these amendments to claim 21 can be found in the specification, for example, at Table 2, which lists SEQ ID NO:12 as a homolog of a known gp120 receptor (GENESEQAAR32188) that binds gp120 (International Patent WO 93/01820). As mentioned above (See Section II D.), SEQ ID NO:12 shares homology with other C-lectin proteins known to bind gp120 and contains a C-type lectin domain that may mediate carbohydrate recognition of viral envelope glycoproteins, including gp120.

The requirements necessary to fulfill the written description requirement of 35 U.S.C. 112, first paragraph, are well established by case law.

the applicant must also convey with reasonable clarity to those skilled in

the art that, as of the filing date sought, he or she was in possession of *the invention*. The invention is, for purposes of the "written description" inquiry, *whatever is now claimed*. *Vas-Cath, Inc. v. Mahurkar*, 19 USPQ2d 1111, 1117 (Fed. Cir. 1991)

Attention is also drawn to the Patent and Trademark Office's own "Guidelines for Examination of Patent Applications Under the 35 U.S.C. Sec. 112, para. 1", published January 5, 2001, which provide that :

An applicant may also show that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics⁴² which provide evidence that applicant was in possession of the claimed invention,⁴³ i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics.⁴⁴ What is conventional or well known to one of ordinary skill in the art need not be disclosed in detail.⁴⁵ If a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described in the specification, then the adequate description requirement is met.⁴⁶

Thus, the written description standard is fulfilled by both what is specifically disclosed and what is conventional or well known to one skilled in the art.

SEQ ID NO:12 and SEQ ID NO:25 are specifically disclosed in the application (see, for example, pages 21-22). Variants of SEQ ID NO:12 and SEQ ID NO:25 are described, for example, at page 22, line 23 through page 23, line 4. Incyte clones in which the nucleic acids encoding the human HCSRPs were first identified and libraries from which those clones were isolated are described, for example, at Table 1 of the Specification. Chemical and structural features of SEQ ID NO:12 are described, for example, in Table 2. Given SEQ ID NO:12 and SEQ ID NO:25, one of ordinary skill in the art would recognize naturally-occurring variants of SEQ ID NO:12 having 90% sequence identity to SEQ ID NO:12 and naturally-occurring variants of SEQ ID NO:25 having 90% sequence identity to SEQ ID NO:25. Accordingly, the Specification provides an adequate written description of the recited polynucleotide and polypeptide sequences.

A. The Specification provides an adequate written description of the claimed "variants" of SEQ ID NO:12 and SEQ ID NO:25.

The Office Action has further asserted that the claims are not supported by an adequate written description because "it cannot be established that a representative number of species have been disclosed to support the genus claim" (Office Action, page 10).

Such a position is believed to present a misapplication of the law.

1. The present claims specifically define the claimed genus through the recitation of chemical structure

Court cases in which "DNA claims" have been at issue (which are hence relevant to claims to proteins encoded by the DNA) commonly emphasize that the recitation of structural features or chemical or physical properties are important factors to consider in a written description analysis of such claims. For example, in *Fiers v. Revel*, 25 USPQ2d 1601, 1606 (Fed. Cir. 1993), the court stated that:

If a conception of a DNA requires a precise definition, such as by structure, formula, chemical name or physical properties, as we have held, then a description also requires that degree of specificity.

In a number of instances in which claims to DNA have been found invalid, the courts have noted that the claims attempted to define the claimed DNA in terms of functional characteristics without any reference to structural features. As set forth by the court in *University of California v. Eli Lilly and Co.*, 43 USPQ2d 1398, 1406 (Fed. Cir. 1997):

In claims to genetic material, however, a generic statement such as "vertebrate insulin cDNA" or "mammalian insulin cDNA," without more, is not an adequate written description of the genus because it does not distinguish the claimed genus from others, except by function.

Thus, the mere recitation of functional characteristics of a DNA, without the definition of structural features, has been a common basis by which courts have found invalid claims to DNA. For example, in *Lilly*, 43 USPQ2d at 1407, the court found invalid for violation of the written description requirement the following claim of U.S. Patent No. 4,652,525:

1. A recombinant plasmid replicable in procarvotic host containing within its nucleotide

sequence a subsequence having the structure of the reverse transcript of an mRNA of a vertebrate, which mRNA encodes insulin.

In *Fiers*, 25 USPQ2d at 1603, the parties were in an interference involving the following count:

A DNA which consists essentially of a DNA which codes for a human fibroblast interferon-beta polypeptide.

Party Revel in the *Fiers* case argued that its foreign priority application contained an adequate written description of the DNA of the count because that application mentioned a potential method for isolating the DNA. The Revel priority application, however, did not have a description of any particular DNA structure corresponding to the DNA of the count. The court therefore found that the Revel priority application lacked an adequate written description of the subject matter of the count.

Thus, in *Lilly* and *Fiers*, nucleic acids were defined on the basis of functional characteristics and were found not to comply with the written description requirement of 35 U.S.C. §112; *i.e.*, "an mRNA of a vertebrate, which mRNA encodes insulin" in *Lilly*, and "DNA which codes for a human fibroblast interferon-beta polypeptide" in *Fiers*. In contrast to the situation in *Lilly* and *Fiers*, the claims at issue in the present application define polynucleotides or polypeptides specifically in terms of chemical structure, rather than on functional characteristics. For example, the "variant language" of independent claims 21 and 30 recite chemical structure to define the claimed genus:

21. An isolated polypeptide selected from the group consisting of: ...
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:12...
30. An isolated polynucleotide selected from the group consisting of: ...
 - b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:25...

From the above it should be apparent that the claims of the subject application are fundamentally different from those found invalid in *Lilly* and *Fiers*. The subject matter of the present claims is defined in terms of the chemical structures of SEQ ID NO:12 and SEQ ID NO:25. In the present case, there is no reliance merely on a description of functional characteristics of the

characteristics. Moreover, if such functional recitations were included, it would add to the structural characterization of the recited polynucleotides or polypeptides. The polynucleotides or polypeptides defined in the claims of the present application recite structural features, and cases such as *Lilly* and *Fiers* stress that the recitation of structure is an important factor to consider in a written description analysis of claims of this type. By failing to base its written description inquiry "on whatever is now claimed," the Office Action failed to provide an appropriate analysis of the present claims and how they differ from those found not to satisfy the written description requirement in *Lilly* and *Fiers*.

2. The present claims do not define a genus which is "highly variant"

Furthermore, the claims at issue do not describe a genus which could be characterized as "highly variant." Available evidence illustrates that the claimed genus is of narrow scope.

In support of this assertion, the Examiner's attention is directed to the enclosed reference by Brenner et al. ("Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships," Proc. Natl. Acad. Sci. USA (1998) 95:6073-6078). Through exhaustive analysis of a data set of proteins with known structural and functional relationships and with <90% overall sequence identity, Brenner et al. have determined that 30% identity is a reliable threshold for establishing evolutionary homology between two sequences aligned over at least 150 residues. (Brenner et al., pages 6073 and 6076.) Furthermore, local identity is particularly important in this case for assessing the significance of the alignments, as Brenner et al. further report that $\geq 40\%$ identity over at least 70 residues is reliable in signifying homology between proteins. (Brenner et al., page 6076.)

The present application is directed, *inter alia*, to human cell surface receptor proteins related to the amino acid sequence of SEQ ID NO:12. In accordance with Brenner et al, naturally occurring molecules may exist which could be characterized as human cell surface receptor proteins and which have as little as 40% identity over at least 70 residues to SEQ ID NO:12. The "variant language" of the present claims recites, for example, polypeptides or polynucleotides encoding "a naturally-occurring amino acid sequence having at least 90% sequence identity to the sequence of SEQ ID NO:12" (note that SEQ ID NO:12 has 325 amino acid residues). This variation is far less than that of all potential human cell surface receptor proteins related to SEQ ID NO:12, i.e., those human cell surface receptor proteins having as little as 40% identity over at least 70 residues to SEQ ID NO:12.

3. The state of the art at the time of the present invention is further advanced than at the time of the *Lilly* and *Fiers* applications

In the *Lilly* case, claims of U.S. Patent No. 4,652,525 were found invalid for failing to comply with the written description requirement of 35 U.S.C. §112. The '525 patent claimed the benefit of priority of two applications, Application Serial No. 801,343 filed May 27, 1977, and Application Serial No. 805,023 filed June 9, 1977. In the *Fiers* case, party Revel claimed the benefit of priority of an Israeli application filed on November 21, 1979. Thus, the written description inquiry in those case was based on the state of the art at essentially at the "dark ages" of recombinant DNA technology.

The present application has a priority date of March 8, 1999. Much has happened in the development of recombinant DNA technology in the 22 or more years from the time of filing of the applications involved in *Lilly* and *Fiers* and the present application. For example, the technique of polymerase chain reaction (PCR) was invented. Highly efficient cloning and DNA sequencing technology has been developed. Large databases of protein and nucleotide sequences have been compiled. Much of the raw material of the human and other genomes has been sequenced. With these remarkable advances one of skill in the art would recognize that, given the sequence information of SEQ ID NO:12 and SEQ ID NO:25, and the additional extensive detail provided by the subject application, the present inventors were in possession of the claimed polynucleotide and polypeptide variants at the time of filing of this application.

4. Summary

The Office Action failed to base its written description inquiry "on whatever is now claimed." Consequently, the Action did not provide an appropriate analysis of the present claims and how they differ from those found not to satisfy the written description requirement in cases such as *Lilly* and *Fiers*. In particular, the claims of the subject application are fundamentally different from those found invalid in *Lilly* and *Fiers*. The subject matter of the present claims is defined in terms of the chemical structure of SEQ ID NO:12 or SEQ ID NO:25. The courts have stressed that structural features are important factors to consider in a written description analysis of claims to nucleic acids and proteins. In addition, the genus of polynucleotides or polypeptides defined by the present claims is adequately described, as evidenced by Brenner et al and consideration of the claims of the '740 patent involved in *Lilly*. Furthermore, there have been remarkable advances in the state of the art since the *Lilly* and

Fiers cases, and these advances were given no consideration whatsoever in the position set forth by the Office Action. Therefore, withdrawal of the rejections under U.S.C. § 112, first paragraph is respectfully requested.

Rejection under 35 U.S.C. § 112, second paragraph

Claims 21-31 and 35-37 have been rejected under 35 U.S.C. § 112, second paragraph, as allegedly being "indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention." (Office Action, page 10). In particular, the Examiner states that it is not clear what is meant by the term "naturally occurring". Applicants respectfully disagree and traverse the rejection.

The M.P.E.P. provides guidelines to Examiners for rejections under 35 U.S.C. § 112, second paragraph as follows:

...a full explanation of the deficiency of the claims should be supplied. Whenever possible, identify the particular term(s) or limitation(s) which render the claim(s) indefinite and state why such term or limitation renders the claim indefinite. If the scope of the claimed subject matter can be determined by one having ordinary skill in the art, a rejection using this form paragraph would not be appropriate.
(M.P.E.P. § 706.03(d))

Therefore claims must be examined on the basis of whether one having ordinary skill in the art would be able to determine the scope of the claim and, if a rejection is made, reasons must be provided why the claim is indefinite. Applicants submit that the Examiner has not provided any reasons or evidence why the cited phrase is indefinite and/or why one having ordinary skill in the art could not determine the scope of the claim. For this reason alone, the rejection is improper and should be withdrawn.

The term "naturally occurring" defines the amino acid and nucleotide sequences encompassed by the claims. "Naturally occurring" means present in or produced by nature. The use of the term "naturally-occurring" distinguishes a sequence that occurs in nature from synthetic or engineered sequences that do not. One of skill in the art would understand the meaning of the term "naturally-occurring" within the context of the claims. Applicants are claiming those polynucleotides and polypeptides comprising sequences that already exist in nature. Applicants wish to clarify that the claim language does not preclude making such sequences synthetically. Claims 21-27, 30, 31, 35, and 36 are composition of matter claims that do not refer to methods of making the polynucleotides and

polypeptides. Applicants have the right to claim a subgenus of variants of SEQ ID NO:12 and SEQ ID NO:25, those comprising naturally occurring sequences, as opposed to all variant sequences that can be made synthetically. The scope of these claims is clear. Therefore, withdrawal of the rejections under U.S.C. § 112, second paragraph is respectfully requested.

Rejections under 35 U.S.C. § 102

Claims 21(d) and 31 have been canceled, thus the rejections under 35 U.S.C. § 102 with respect to these claims and their dependent claims are moot.

CONCLUSION

In light of the above amendments and remarks, Applicants submit that the present application is fully in condition for allowance, and request that the Examiner withdraw the outstanding rejections. Early notice to that effect is earnestly solicited.

If the Examiner contemplates other action, or if a telephone conference would expedite allowance of the claims, Applicants invite the Examiner to contact Applicants' Attorney at (650) 855-0555.

Applicants believe that no fee is due with this communication. However, if the USPTO determines that a fee is due, the Commissioner is hereby authorized to charge Deposit Account No. 09-0108.

Respectfully submitted,
INCYTE GENOMICS, INC.

Date: 15 April 2003

Jenny Buchbinder
Jenny Buchbinder
Reg. No. 48,588
Direct Dial Telephone: (650) 843-7212

Date: 15 April 2003

Cathleen M. Rocco
Cathleen M. Rocco
Reg. No. 46,172
Direct Dial Telephone: (650) 845-4587

3160 Porter Drive
Palo Alto, California 94304
Phone: (650) 855-0555
Fax: (650) 849-8886

VERSION WITH MARKINGS TO SHOW CHANGES MADE

IN THE SPECIFICATION:

The following sentence was added immediately after the title of the application on page 1 of the specification:

This application claims benefit under 35 U.S.C. § 119(e) of provisional application 60/183,048, filed on November 12, 1999, provisional application 60/183,045 filed on December 7, 1998, and provisional application 60/123,404 filed on March 8, 1999.

The sixth column of Table 2 at page 63 of the application, beginning with the row for SEQ ID NO:12, was replaced with the following rewritten column:

Homologous Sequences
Non-CD4 glycoprotein gp120 receptor <u>GENESEQ AAR32188</u>
DP prostanoid receptor g940379

IN THE CLAIMS:

Claim 31 has been canceled.

Claims 21, 22, 25, 29, 30, and 36 have been amended as follows:

21. (Once Amended) An isolated polypeptide selected from the group consisting of:

- a) a polypeptide comprising an amino acid sequence [selected from the group consisting of SEQ ID NO:1-13] of SEQ ID NO:12,
- b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to [an] the amino acid sequence [selected from the group consisting of SEQ ID NO:1-13] of SEQ ID NO:12, wherein said polypeptide binds to human immunodeficiency virus glycoprotein gp120, and
- c) a biologically active fragment of a polypeptide having [an] the amino acid sequence [selected from the group consisting of SEQ ID NO:1-13] of SEQ ID NO:12, wherein said fragment binds to human immunodeficiency virus glycoprotein gp120. [, and
- d) an immunogenic fragment of a polypeptide having [an] the amino acid sequence [selected from the group consisting of SEQ ID NO:1-13] of SEQ ID NO:12.]

22. (Once Amended) An isolated polypeptide of claim 21 comprising [an] the amino acid sequence [selected from the group consisting of SEQ ID NO:1-13] of SEQ ID NO:12.

25. (Once Amended) An isolated polynucleotide of claim 24 comprising a polynucleotide sequence [selected from the group consisting of SEQ ID NO:14-26] of SEQ ID NO:25.

29. (Once Amended) A method of claim 28, wherein the polypeptide comprises [an] the amino acid sequence [selected from the group consisting of SEQ ID NO:1-13] of SEQ ID NO:12.

30. (Once Amended) An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence [selected from the group consisting of SEQ ID NO:14-26] of SEQ ID NO:25,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to [a] the polynucleotide sequence [selected from the group consisting of SEQ ID NO:14-26] of SEQ ID NO:25,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).

36. (Once Amended) A composition of claim 35, wherein the polypeptide comprises [an] the amino acid sequence [selected from the group consisting of SEQ ID NO:1-13] of SEQ ID NO:12.

Certificate of Mailing

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: **Box Non-Fee Amendment**, Commissioner for Patents, Washington, D.C. 20231 on _____.

By: _____ Printed: _____

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

RECEIVED

In re Application of: Tang et al.

APR 22 2003

Title: HUMAN CELL SURFACE RECEPTORS

TECH CENTER 1600/2900

Serial No.: 09/831,458

Filing Date: May 08, 2001

Examiner: O'Hara, E.

Group Art Unit: 1646

RECEIVED

Box Non-Fee Amendment
Commissioner for Patents
Washington, D.C. 20231

JUL 09 2003

TECH CENTER 1600/2900

**DECLARATION OF DR. TOD BEDILION
UNDER 37 C.F.R. § 1.132**

I, TOD BEDILION, a citizen of the United States, residing at 132 Winding Way, San Carlos, California, declare that:

1. I was employed by Incyte Genomics, Inc. (hereinafter "Incyte") as a Director of Corporate Development until May 11, 2001. I am currently under contract to be a Consultant to Incyte Genomics, Inc.

2. In 1996, I received a Ph.D. degree in Cell, Molecular and Development Biology from UCLA. I had previously received, in 1988, a B.S. degree in biology from UCLA.

Upon my graduation from UCLA, I became, in April 1996, the first employee of Synteni, Inc. (hereinafter "Synteni"). I was a Research Director at Synteni from April 1996 until Synteni was acquired by Incyte in early 1998.

I understand that Synteni was founded in 1994 by T. Dari Shalon while he was a graduate student at Stanford University. I further understand that Synteni was founded for the purpose

of commercially exploiting certain "cDNA microarray" technology that was being worked on at Stanford in the early to mid-1990s. That technology, which I will sometimes refer to herein as the "Stanford-developed cDNA microarray technology", was the subject of Dr. Shalon's doctoral thesis at Stanford. I understand and believe that Dr. P.O. Brown was Dr. Shalon's thesis advisor at Stanford.

During the period beginning before I was employed by Synteni and ending upon its acquisition by Incyte in early 1998, I understand Synteni was the exclusive licensee of the Stanford-developed cDNA microarray technology, subject to any right that the United States government may have with respect to that technology. In early 1998, I understand Incyte acquired rights under the Stanford-developed cDNA microarray technology as part of its acquisition of Synteni.

I understand that at the time of the commencement of my employment at Synteni in April 1996, Synteni's rights with respect to the Stanford-developed cDNA technology included rights under a United States patent application that had been filed June 7, 1995 in the names of Drs. Brown and Shalon and that subsequently issued as United States Patent No. 5,807,522 (the Brown '522 patent). In December 1995, the subject matter of the Brown '522 patent was published based on a PCT patent application that had also been filed in June 1995. The Brown '522 patent (and its corresponding PCT application) describes the use of the Stanford-developed cDNA technology in a number of gene expression monitoring applications, as will be discussed more fully below.

Upon Incyte's acquisition of Synteni, I became employed by Incyte. From early 1998 until late 1999, I was an Associate Research Director at Incyte. In late 1999, I was promoted to the position of Director, Corporate Development.

I have been aware of the Stanford-developed cDNA microarray technology since shortly before I commenced my employment at Synteni. While I was employed by Synteni, virtually all (if not all) of my work efforts (as well as the work efforts of others employed by Synteni) were directed to the further development and commercial exploitation of that cDNA microarray technology. By the

end of 1997, those efforts had progressed to the point that I understand Incyte agreed to pay at least about \$80 million to acquire Synteni. Since I have been employed by Incyte, I have continued to work

on the further development and commercial exploitation of the cDNA microarray technology that was first developed at Stanford in the early to mid-1990s.

3. I have reviewed the specification of a United States patent application that I understand was filed on May 8, 2001 in the names of Tang et al. and was assigned Serial No. 09/831,458 (hereinafter "the Tang et al. '458 application"). Furthermore, I understand that this United States patent application claimed priority to United States Provisional Patent Application Serial No. 60/123,404 filed on March 8, 1999 (hereinafter "the Tang et al. '404 application"). The SEQ ID NO:12-encoding polynucleotides were described in the Tang et al. '404 application. (Note that the sequences of SEQ ID NO:12 and SEQ ID NO:25 disclosed in the Tang et al. '458 application are identical to the sequences referred to as SEQ ID NO:5 and SEQ ID NO:11, respectively, in the Tang et al. '404 application). My remarks herein will therefore be directed to the Tang et al. '404 patent application, and March 8, 1999, as the relevant date of filing. In broad overview, the Tang et al. '404 specification pertains to certain nucleotide and amino acid sequences and their use in a number of applications, including gene expression monitoring applications that are useful in connection with (a) developing drugs (e.g., the diagnosis of inherited and acquired genetic disorders, expression profiling, toxicology testing, and drug development with respect to cancer, an immunopathology, a neuropathology, and the like), and (b) monitoring the activity of drugs for purposes relating to evaluating their efficacy and toxicity.

4. I understand that (a) the Tang et al. '404 application contains claims that are directed to isolated and purified polynucleotides having the sequences disclosed in the Tang et al. '458 application as SEQ ID NO:12-encoding polynucleotides, for example SEQ ID NO:25 (hereinafter "the SEQ ID NO:12-encoding polynucleotides"), and (b) the Patent Examiner has rejected those claims on

the grounds that the specification of the Tang et al. '404 application does not disclose a substantial, specific and credible utility for the claimed SEQ ID NO:12-encoding polynucleotides. I further understand that whether or not a patent specification discloses a substantial, specific and credible utility for its claimed subject matter is properly determined from the perspective of a person skilled in the art to which the specification pertains at the time of the patent application was filed. In addition, I understand that a substantial, specific and credible utility under the patent laws must be a "real-world" utility.

5. I have been asked (a) to consider with a view to reaching a conclusion (or conclusions) as to whether or not I agree with the Patent Examiner's position that the Tang et al. '404 application does not disclose a substantial, specific and credible "real-world" utility for the claimed SEQ ID NO:12-encoding polynucleotides, and (b) to state and explain the bases for any conclusions I reach. I have been informed that, in connection with my considerations, I should determine whether or not a person skilled in the art to which the Tang et al. '404 application pertains on March 8, 1999 would have concluded that the Tang et al. '404 application disclosed, for the benefit of the public, a specific beneficial use of the SEQ ID NO:12-encoding polynucleotides in their then available and disclosed form. I have also been informed that, with respect to the "real-world" utility requirement, the Patent and Trademark Office instructs its Patent Examiners in Section 2107 of the Manual of Patent Examining Procedure, under the heading "I. 'Real-World Value' Requirement":

"Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific and unquestionable utility (e.g., they are useful in analyzing compounds). An assessment that focuses on whether an invention is useful only in a research setting thus does not address whether the specific invention is in fact 'useful' in a patent sense. Instead, Office personnel must distinguish between inventions that have a specifically identified utility and inventions whose specific utility requires further research to identify or reasonably confirm."

6. I have considered the matters set forth in paragraph 5 of this Declaration and have concluded that, contrary to the position I understand the Patent Examiner has taken, the specification of the Tang et al. '404 application does disclose a substantial, specific and credible

of its filing a number of substantial, specific and credible real-world utilities for the claimed SEQ ID NO:12-encoding polynucleotides. More specifically, persons skilled in the art on March 8, 1999 would have understood the Tang et al. '404 application to disclose the use of the SEQ ID NO:12-encoding polynucleotides in a number of gene expression monitoring applications that were well-known at that time to be useful in connection with the development of drugs and the monitoring of the activity of such drugs. I explain the bases for reaching my conclusion in this regard in paragraphs 7-16 below.

7. In reaching the conclusion stated in paragraph 6 of this Declaration, I considered (a) the specification of the Tang et al. '404 application, and (b) a number of published articles and patent documents that evidence gene expression monitoring techniques that were well-known before the March 8, 1999 filing date of the Tang et al. '404 application. The published articles and patent documents I considered are:

(a) Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W., Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes, Proc. Natl. Acad. Sci. USA, 93, 10614-10619 (1996) (hereinafter "the Schena 1996 article") (copy annexed at Tab A);

(b) Schena, M., Shalon, D., Davis, R.W., Brown, P.O., Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray, Science, 270, 467-470 (1995) (hereinafter "the Schena 1995 article") (copy annexed at Tab B);

(c) Shalon and Brown PCT patent application WO 95/35505 titled "Method and Apparatus For Fabricating Microarrays Of Biological Samples," filed on June 16, 1995, and published on December 28, 1995 (hereinafter "the Shalon PCT application") (copy annexed at Tab C);

(d) Brown and Shalon U.S. Patent No. 5,807,522, corresponding to the Shalon PCT application, titled "Methods For Fabricating Microarrays Of Biological Samples," filed on June 7, 1995 and issued on September 15, 1998 (hereinafter "the Brown '522 patent") (copy annexed at Tab D);

(e) DeRisi, J., Penland, L., and Brown, P.O. (Group 1); Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., and Trent, J.M. (Group 2), Use of a cDNA microarray to analyse gene expression patterns in human cancer, Nat. Genet., 14(4), 457-460 (1996) (hereinafter "the DeRisi article") (copy annexed at Tab E);

(f) Shalon, D., Smith, S.J., and Brown, P.O., A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization, Genome Res., 6(7), 639-645 (1996) (hereinafter "the Shalon article") (copy annexed at Tab F);

(g) Heller, R.A., Schena, M., Chai A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E., and Davis R.W., Discovery and analysis of inflammatory disease-related genes using cDNA microarrays, Proc. Natl. Acad. Sci. USA, 94, 2150-2155 (1997) (hereinafter "the Heller article")(copy annexed at Tab G);

(h) Sambrook, J., Fritsch, E.F., Maniatis, T., Molecular Cloning, A Laboratory Manual, pages 7.37 and 7.38, Cold Spring Harbor Press (1989) (hereinafter "the Sambrook Manual") (copy annexed at Tab H);

8. Many of the published articles and patent documents I considered (i.e., at least items (a)-(g) identified in paragraph 7) relate to work done at Stanford University in the early and mid-1990s with respect to the development of cDNA microarrays for use in gene expression monitoring applications under which Synteni became exclusively licensed. As I will discuss, a person skilled in the art who read the Tang et al. '404 application on March 8, 1999 would have understood that application to disclose the SEQ ID NO:12-encoding polynucleotides to be useful for a number of gene expression monitoring applications, e.g., as a probe for the expression of that specific polynucleotide in cDNA microarrays of the type first developed at Stanford.

9. Turning more specifically to the Tang et al. '404 specification, the SEQ ID NO:25 polynucleotide (referred to as SEQ ID NO:11 in the Tang et al. '404 application) is shown at p. 11 as one of 12 sequences under the heading "Sequence Listing." The Tang et al. '404 specification specifically teaches that the "the invention features isolated and substantially purified polynucleotides that encode HRP" (Tang et al. '404 application at p. 4). It further teaches that (a) the identity of the

SEQ ID NO:25 polynucleotide was determined from a spleen tissue cDNA library (SPLNNOT09) (Tang et al. '404 application, Table 1), (b) the SEQ ID NO:25 polynucleotide encodes for the human receptor protein (HRP) shown as SEQ ID NO:12 (Tang et al. '404 application at Table 2), and (c) northern analysis of SEQ ID NO:25 shows its expression predominantly in cDNA libraries associated with hematopoietic/immune, gastrointestinal, cardiovascular, and reproductive tissues (Tang et al. '404 application at Table 3).

The Tang et al. '404 application discusses a number of uses of the SEQ ID NO:12-encoding polynucleotides in addition to their use in gene expression monitoring applications. I have not fully evaluated these additional uses in connection with the preparation of this Declaration and do not express any views in this Declaration regarding whether or not the Tang et al. '404 specification discloses these additional uses to be substantial, specific and credible real-world utilities of the SEQ ID NO:12-encoding polynucleotides. Consequently, my discussion in this Declaration concerning the Tang et al. '404 application focuses on the portions of the application that relate to the use of the SEQ ID NO:12-encoding polynucleotides in gene expression monitoring applications.

10. The Tang '404 application discloses that the polynucleotide sequences disclosed therein, including the SEQ ID NO:12-encoding polynucleotides, are useful as probes in microarrays. It further teaches that the microarrays can be used "to monitor the expression level of large numbers of genes simultaneously" for a number of purposes, including "to develop and monitor the activities of therapeutic agents" (Tang '404 application at p. 33, lines 18-22).

In the paragraph immediately following the Tang et al. '404 teachings described in the preceding paragraph of this Declaration, the Tang '404 application teaches that microarrays can be prepared using the previously mentioned cDNA microarray technology developed at Stanford in the early to mid-1990s. In this connection, the Tang '404 application specifically cites to the Schena 1996 article identified in item (a) of paragraph 7 of this Declaration (Tang '404 application at p. 32; *supra*, paragraph 7).

The Schena 1996 article is one of a number of documents that were published prior to the March 8, 1999 filing date of the Tang et al. '404 application that describes the use of the Stanford-

developed cDNA technology in a wide range of gene expression monitoring applications, including monitoring and analyzing gene expression patterns in human cancer. In view of the Tang et al. '404 application, the Schena 1996 article, and other related pre-March, 1999 publications, persons skilled in the art on March 8, 1999 clearly would have understood the Tang et al. '404 application to disclose the SEQ ID NO:12-encoding polynucleotides to be useful in cDNA microarrays for the development of new drugs and monitoring the activities of drugs for such purposes as evaluating their efficacy and toxicity, as explained more fully in paragraph 15 below.

With specific reference to toxicity evaluations, those of skill in the art who were working on drug development in March, 1999 (and for many years prior to March, 1999) without any doubt appreciated that the toxicity (or lack of toxicity) of any proposed drug they were working on was one of the most important criteria to be considered and evaluated in connection with the development of the drug. They would have understood at that time that good drugs are not only potent, they are specific. This means that they have strong effects on a specific biological target and minimal effects on all other biological targets. Ascertaining that a candidate drug affects its intended target, and identification of undesirable secondary effects (i.e., toxic side effects), had been for many years among the main challenges in developing new drugs. The ability to determine which genes are positively affected by a given drug, coupled with the ability to quickly and at the earliest time possible in the drug development process identify drugs that are likely to be toxic because of their undesirable secondary effects, have enormous value in improving the efficiency of the drug discovery process, and are an important and essential part of the development of any new drug. Accordingly, the teachings in the Tang et al. '404 application, in particular regarding use of the SEQ ID NO:12-encoding polynucleotides in differential gene expression analysis and in the development and the monitoring of the activities of drugs, clearly includes toxicity studies and persons skilled in the art who read the Tang et al. '404 application on March 8, 1999 would have understood that to be so.

11. The Schena 1996 article was not the first publication that described the use of the cDNA microarray technique developed at Stanford to monitor quantitatively gene expression patterns. More than a year earlier (i.e., in October 1995), the Schena 1995 article, titled "Quantitative

Monitoring of Gene Expression Patterns with a Complementary DNA Microarray", was published (see Tabs A and B).

12. As previously discussed (*supra*, paragraphs 2 and 7), in the mid-1990s patent applications were filed in the names of Drs. Shalon and Brown that described the Stanford-developed cDNA microarray technology. The two patent documents (i.e., the Shalon PCT application and the Brown '522 patent) annexed to this Declaration at Tabs C and D evidence information that was available to the public regarding the Stanford-developed cDNA microarray technology before the March 8, 1999 filing date of the Tang et al. '404 application.

The Shalon PCT patent application, which was published in December 1995, contains virtually the same (if not exactly the same) specification as the Brown '522 patent. Hence, the Brown '522 patent disclosure was, in effect, available to the public as of the December 1995 publication date of the Shalon PCT application (see Tabs C and D). For the sake of convenience, I cite to and discuss the Brown '522 specification below on the understanding that the descriptions in that specification were published as of the December 28, 1995 publication date of the Shalon PCT application.

The Brown '522 patent discusses, in detail, the utility of the Stanford-developed cDNA microarrays in gene expression monitoring applications. For example, in the "Summary Of The Invention" section, the Brown '522 patent teaches (see Tab D, col. 4, line 52-col. 5, line 8):

Also forming part of the invention is a method of detecting differential expression of each of a plurality of genes in a first cell type, with respect to expression of the same genes in a second cell type. In practicing the method, there is first produced fluorescent-labeled cDNAs from mRNAs isolated from two cells types, where the cDNAs from the first and second cell types are labeled with first and second different fluorescent reporters.

A mixture of the labeled cDNAs from the two cell types is added to an array of polynucleotides representing a plurality of known genes derived from the two cell types, under conditions that result in hybridization of the cDNAs to complementary-sequence polynucleotides in the array. The array is then examined by fluorescence under fluorescence excitation conditions in which (i) polynucleotides in the array that are hybridized predominantly to

cDNAs derived from one of the first or second cell types give a distinct first and second fluorescence emission color, respectively, and (ii) polynucleotides in the array that are hybridized to substantially equal numbers of cDNAs derived from the first and second cell types give a distinct combined fluorescence emission color, respectively. The relative expression of known genes in the two cell types can then be determined by the observed fluorescence emission color of each spot.

The Brown '522 patent further teaches that the “[m]icroarrays of immobilized nucleic acid sequences prepared in accordance with the invention” can be used in “numerous” genetic applications, including “monitoring of gene expression” applications (see Tab D at col. 14, lines 36-42). The Brown '522 patent teaches (a) monitoring gene expression (i) in different tissue types, (ii) in different disease states, and (iii) in response to different drugs, and (b) that arrays disclosed therein may be used in toxicology studies (see Tab D at col. 15, lines 13-18 and 52-58 and col. 18, lines 25-30).

13. Also pertinent to my considerations underlying this Declaration is the DeRisi article, published in December 1996. The DeRisi article describes the use of the Stanford-developed cDNA microarray technology “to analyze gene expression patterns in human cancer” (see Tab E at, e.g., p. 457). The DeRisi article specifically indicates, consistent with what was apparent to persons skilled in the art in December 1996, that increasing the number of genes on the cDNA microarray permits a “more comprehensive survey of gene expression patterns,” thereby enhancing the ability of the cDNA microarray to provide “new and useful insights into human biology and a deeper understanding of the gene pathways involved in the pathogenesis of cancer and other diseases” (see Tab E at p. 458).

14. Other pre-March, 1999 publications further evidence the utility of the cDNA microarrays first developed at Stanford in a wide range of gene expression monitoring applications (see, e.g., the Shalon and the Heller articles at Tabs F and G). By no later than the March 1997 publication of the Heller article, these publications showed that employees of Synteni (i.e., James Gilmore and myself) had used the cDNA microarrays in specific gene expression monitoring applications (see Tab G).

The Heller article states that the results reported therein "successfully demonstrate the use of the cDNA microarray system as a general approach for dissecting human diseases" (Tab G at p. 2150). Among other things, the Heller article describes the investigation of "1000 human genes that were randomly selected from a peripheral human blood cell library" and "[t]heir differential and quantitative expression analysis in cells of the joint tissue. . . to demonstrate the utility of the microarray method to analyze complex diseases by their pattern of gene expression" (see Tab G at pp. 2150 *et seq.*).

Much of the work reported on in the Heller article was done in 1996. That article, therefore, evidences how persons skilled in the art were readily able, well prior to March 8, 1999, to make and use cDNA microarrays to achieve highly useful results. For example, as reported in the Heller article, a cDNA microarray that was used in some of the highly successful work reported on therein was made from 1,000 genes randomly selected from a human blood cell library.

15. A person skilled in the art on March 8, 1999, who read the Tang et al. '404 application, would understand that application to disclose the SEQ ID NO:12-encoding polynucleotides, for example, SEQ ID NO:25, to be highly useful as probes for the expression of that specific polynucleotide in cDNA microarrays of the type first developed at Stanford. For example, the specification of the Tang et al. '404 application would have led a person skilled in the art in March, 1999 who was using gene expression monitoring in connection with working on developing new drugs for the treatment of cell proliferative disorders, immune system disorders, infections, and neuronal disorders to conclude that a cDNA microarray that contained the SEQ ID NO:12-encoding polynucleotides would be a highly useful tool and to request specifically that any cDNA microarray that was being used for such purposes contain the SEQ ID NO:12-encoding polynucleotides. Persons skilled in the art would appreciate that cDNA microarrays that contained the SEQ ID NO:12-encoding polynucleotides would be a more useful tool than cDNA microarrays that did not contain the polynucleotides in connection with conducting gene expression monitoring studies on proposed (or actual) drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders for such purposes as evaluating their efficacy and toxicity.

I discuss in more detail in items (a)-(g) below a number of reasons why a person skilled in the art, who read the Tang et al. '404 specification in March, 1999, would have concluded based on that specification and the state of the art at that time, that the SEQ ID NO:12-encoding polynucleotides would be a highly useful tool for inclusion in cDNA microarrays for evaluating the efficacy and toxicity of proposed drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders , as well as for other evaluations:

(a) The Tang et al. '404 application teaches the SEQ ID NO:12-encoding polynucleotides to be useful as probes in cDNA microarrays of the type first developed at Stanford. It also teaches that such cDNA microarrays are useful in a number of gene expression monitoring applications, including "developing and monitoring the activity of therapeutic agents [i.e., drugs]" (see paragraph 10, *supra*).

(b) By March, 1999, the Stanford-developed cDNA microarray technology was a well known and widely accepted tool for use in a wide range of gene expression monitoring applications. This is evidenced, for example, by numerous publications describing the use of that cDNA technology in gene expression monitoring applications and the fact that, for over a year, the technology had provided the basis for the operations of an up-and-running company (Synteni), with employees, that was created for the purpose of developing and commercially exploiting that technology (see paragraphs 2, 8 and 10-14, *supra*). The fact that Incyte agreed to purchase Synteni in late 1997 for an amount reported to be at least about \$80 million only serves to underscore the substantial practical and commercial significance, in 1997, of the cDNA microarray technology first developed at Stanford (see paragraph 2, *supra*).

(c) The pre-March, 1999 publications regarding the cDNA microarray technology first developed at Stanford that I discuss in this Declaration repeatedly confirm that, consistent with the teachings in the Tang et al. '404 application, cDNA microarrays are highly useful tools for conducting gene expression monitoring applications with respect to the development of drugs and the monitoring of their activity. Among other things, those pre-March, 1999 publications confirmed that cDNA microarrays (i) were useful for monitoring gene expression responses to different drugs (see paragraph 12, *supra*), (ii) were useful in analyzing gene expression patterns in human cancer, with

increasing the number of genes on the cDNA microarray enhancing the ability of the cDNA microarray to provide useful information (see paragraph 13, *supra*), and (iii) were a valuable tool for use as part of a "general approach for dissecting human diseases" and for "analyz[ing] complex diseases by their pattern of gene expression" (see paragraph 14, *supra*).

(d) Based on my own extensive work for a company whose business was the development and commercial exploitation of cDNA microarray technology for more than two years prior to the March, 1999 filing date of the Tang et al. '404 application, I have first-hand knowledge concerning the state of the art with respect to making and using cDNA microarrays as of March 8, 1999 (see paragraphs 2 and 14, *supra*). Persons skilled in the art as of that date would have (a) concluded that the Tang et al. '404 application disclosed cDNA microarrays containing the SEQ ID NO:12-encoding polynucleotides to be useful, and (b) readily been able to make and use such microarrays with useful results.

(e) The Tang et al. '404 specification contains a number of teachings that would lead persons skilled in the art on March 8, 1999 to conclude that a cDNA microarray that contained the SEQ ID NO:12-encoding polynucleotides would be a more useful tool for gene expression monitoring applications relating to drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders than a cDNA microarray that did not contain the SEQ ID NO:12-encoding polynucleotides. Among other things, the Tang et al. '404 specification teaches that the identity of the SEQ ID NO:25 polynucleotide was determined from a spleen tissue cDNA library (SPLNNOT09) (Tang et al. '404 application, Table 1). Moreover, northern analysis of SEQ ID NO:25 shows its expression predominantly in cDNA libraries associated with hematopoietic/immune, gastrointestinal, cardiovascular, and reproductive tissues. (Tang et al. '404 application at Table 3). (See paragraph 9, *supra*).

Moreover, the Tang et al. '404 specification teaches that the HRP protein having the amino acid sequence of SEQ ID NO:12 shares homology with known functional proteins. HRP is a member of the human receptor protein family. In particular, SEQ ID NO:12 shares homology with the sequence of the non-CD4 glycoprotein gp120 receptor (GENESEQ AAR32188). (Tang et al. '404 application, at Table 2).

(f) Persons skilled in the art on March 8, 1999 would have appreciated (i) that the gene expression monitoring results obtained using a cDNA microarray containing a probe to a sequence selected from the group consisting of SEQ ID NO:12-encoding polynucleotides would vary, depending on the particular drug being evaluated, and (ii) that such varying results would occur both with respect to the results obtained from the probe described in (i) and from the cDNA microarray as a whole (including all its other individual probes). These kinds of varying results, depending on the identity of the drug being tested, in no way detracts from my conclusion that persons skilled in the art on March 8, 1999, having read the Tang et al. '404 specification, would specifically request that any cDNA microarray that was being used for conducting gene expression monitoring studies on drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders (*e.g.*, a toxicology study or any efficacy study of the type that typically takes place in connection with the development of a drug) contain any one of the SEQ ID NO:12 encoding polynucleotides as a probe. Persons skilled in the art on March 8, 1999 would have wanted their cDNA microarray to have a probe as described in (i) because a microarray that contained such a probe (as compared to one that did not) would provide more useful results in the kind of gene expression monitoring studies using cDNA microarrays that persons skilled in the art have been doing since well prior to March 8, 1999.

The foregoing is not intended to be an all-inclusive explanation of all my reasons for reaching the conclusions stated in this paragraph 15, and in paragraph 6, *supra*. In my view, however, it provides more than sufficient reasons to justify my conclusions stated in paragraph 6 of this Declaration regarding the Tang et al. '404 application disclosing to persons skilled in the art at the time of its filing substantial, specific and credible real-world utilities for the SEQ ID NO:12-encoding polynucleotides.

16. Also pertinent to my considerations underlying this Declaration is the fact that the Tang et al. '404 disclosure regarding the uses of the SEQ ID NO:25 polynucleotide for gene expression monitoring applications is not limited to the use of that polynucleotide as a probe in microarrays. For one thing, the Tang et al. '404 disclosure regarding the hybridization technique used

in gene expression monitoring applications is broad (Tang et al. '404 application at, e.g., p. 4, lines 3-8).

In addition, the Tang et al. '404 specification repeatedly teaches that the polynucleotides described therein (including the polynucleotide of SEQ ID NO:25) may desirably be used as probes in any of a number of long established "standard" non-microarray techniques, such as Northern analysis, for conducting gene expression monitoring studies. See, e.g.:

(a) Tang et al. '404 application at p. 4, lines 27-29 ("[N]orthern analysis is indicative of the presence of nucleic acids encoding HRP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding HRP");

(b) Tang et al. '404 application at p. 32, lines 7-11 ("The polynucleotide sequences encoding HRP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered HRP expression. Such qualitative or quantitative methods are well known in the art");

(c) Tang et al. '404 application at p. 32, lines 21-29 ("In order to provide a basis for the diagnosis of a disorder associated with expression of HRP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HRP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder"); and

(d) Tang et al. '404 application at p. 37, lines 18-21 ("Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)").

The "Sambrook et al." reference cited in item (d) immediately above is a reference that was well known to persons skilled in the art in March, 1999. A copy of pages from that reference

manual, which was published in 1989, is annexed to this Declaration at Tab H. The attached pages from the Sambrook manual provide an overview of northern analysis and other membrane-based technologies for conducting gene expression monitoring studies that were known and used by persons skilled in the art for many years prior to the March 8, 1999 filing date of the Tang et al. '404 application.

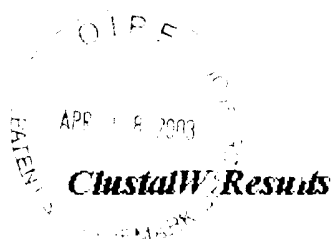
A person skilled in the art on March 8, 1999, who read the Tang et al. '404 specification, would have routinely and readily appreciated that the SEQ ID NO:12-encoding polynucleotides disclosed therein would be useful as a probe to conduct gene expression monitoring analyses using northern analysis or any of the other traditional membrane-based gene expression monitoring techniques that were known and in common use many years prior to the filing of the Tang et al. '404 application. For example, a person skilled in the art in March, 1999 would have routinely and readily appreciated that the SEQ ID NO:12-encoding polynucleotides would be a useful tool in conducting gene expression analyses, using the northern analysis technique, in furtherance of (a) the development of drugs for the treatment of cell proliferative disorders, immune system disorders, infections, and neuronal disorders, and (b) analyses of the efficacy and toxicity of such drugs.

17. I declare further that all statements made herein of my own knowledge are true and that all statements made herein on information and belief are believed to be true; and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, and that willful false statements may jeopardize the validity of this application and any patent issuing thereon.

Tod Bedilion

Signed at Redwood City, California

this ____ day of April, 2003



RECEIVED

JUL 09 2003

TECH CENTER 1600/2900

Sequences

Help

Retrieval

BLAST2

FASTA

ClustalW

GCG Assembly

Phrap

Translation

Confidential

Property of Incyte Genomics, Inc.

SeqServer Version 4.6 Jan 2002

☐ g13383470
☐ g4586836
☐ SEQIDNO_12

RECEIVED

APR 22 2003

TECH CENTER 1600/2900

CLUSTAL W (1.7) Multiple Sequence Alignments

Sequence format is Pearson

Sequence 1: g13383470 376 aa

Sequence 2: g4586836 296 aa

Sequence 3: SEQIDNO_12 325 aa

Start of Pairwise alignments

Aligning...

Sequences (1:2) Aligned. Score: 88

Sequences (1:3) Aligned. Score: 99

Sequences (2:3) Aligned. Score: 80

Start of Multiple Alignment

There are 2 groups

Aligning...

Group 1: Sequences: 2 Score:3450

Group 2: Sequences: 3 Score:3741

Alignment Score 4852

CLUSTAL-Alignment file created [baaWNaqjO.aln]

CLUSTAL W (1.7) multiple sequence alignment

```
g13383470      MSDSKEPRVQQLGLLLEEDPTTSGIRLFPRDFQFQQIHGHKSSTGCLGHGALVLDLLSFML
g4586836      MSDSKEPRVQQLGLL-----GCLGHGALVLDLLSFML
SEQIDNO_12    MSDSKEPRVQQLGLL-----GCLGHGALVLDLLSFML
*****
```

```
g13383470      LAGVLVAILVQVSKVPSSLSQEQQSEQDAIYQNLITQLKAAVGELSEKSKLQEIYQELITQLK
g4586836      LAGVLVAILVQVSKVPSSLSQEQQSEQDAIYQNLITQLKAAVGELSEKSKLQEIYQELITQLK
SEQIDNO_12    LAGVLVAILVQVSKVPSSLSQEQQSEQDAIYQNLITQLKAAVGELSEKSKLQEIYQELITQLK
*****
```

```
g13383470      AAVGELPEKSKLQEIYQELITRLKAAVGELPEKSKLQEIYQELITRLKAAVGELPEKS----
g4586836      AAVGELPEKSKLQEIYQELITRLKAAVGELPEKSKLQEIYQELITRLKAAVGELPEKS-KLQ
SEQIDNO_12    AAVGELPEKSKLQEIYQELITRLKAAVGELPEKSKLQEIYQELITRLKAAVGELPEKS---
*****
```

*****_*****

```
g13383470      ELPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFFQGNCYFMSNSQRINWHDSVTACQEV
g4586836      ELPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFFQGNCYFMSNSQRINWHDSVTACQEV
SEQIDNO_12    ---DQSKQQQIYQELTDLKTAFERLCRHCPKDWTFFQGNCYFMSNSQRINWHDSVTACQEV
               *****
```

```
g13383470      RAQLVVIKTAEEQNFLQLQTSRSNRFSWMGLSDLNQEGTWQWVDGSPLSPSFQRYWNSGE
g4586836      RAQLVVIKTAEEQLPVLEQWRTQQ-----
SEQIDNO_12    RAQLVVIKTAEEQNFLQLQTSRSNRFSWMGLSDLNQEGTWQWVDGSPLSPSFQRYWNSGE
               *****      *:  *:::
```

```
g13383470      PNNSGNEDCAEFSGSGWNDNRCDVDNYWICKKPAACFRDE
g4586836      -----
SEQIDNO_12    PNNSGNEDCAEFSGSGWNDNRCDVDNYWICKKPAACFRDE
```

Submit sequences to:



SeqServer[®]
biology in silico

BLAST2 Search Results

[Sequences](#)[Help](#)[Retrieval](#)[BLAST2](#)[FASTA](#)[ClustalW](#)[GCG Assembly](#)[Phrap](#)[Translation](#)[BLAST2 Manual](#)

Confidential Property of Incyte Genomics, Inc. SeqServer Version 4.6 Jan 2002

Program: blastp

Sequence ID(s):

☐ 3344986CD1 vs. genpept131

NCBI-BLASTP 2.0.10 [Aug-26-1999]



Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= 3344986CD1
(325 letters)

Database: genpept131
1,135,942 sequences; 348,344,575 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
<input checked="" type="checkbox"/> <u>g13383470</u> L-SIGN [Homo sapiens]	635	0.0
<input checked="" type="checkbox"/> <u>g15383606</u> mDC-SIGN2 type I isoform [Homo sapiens]	621	e-177
<input checked="" type="checkbox"/> <u>g12084795</u> probable mannose binding C-type lectin DC-SIGNR [Ho	621	e-177
<input checked="" type="checkbox"/> <u>g12084797</u> probable mannose binding C-type lectin DC-SIGNR [Ho	617	e-175
<input checked="" type="checkbox"/> <u>g15383614</u> sDC-SIGN2 type I isoform [Homo sapiens]	583	e-165
<input checked="" type="checkbox"/> <u>g8572543</u> membrane-associated lectin type-C [Homo sapiens]	532	e-150
<input checked="" type="checkbox"/> <u>g17049084</u> unnamed protein product [Homo sapiens]	532	e-150
<input checked="" type="checkbox"/> <u>g15281073</u> mDC SIGN1A type I isoform [Homo sapiens]	532	e-150
<input checked="" type="checkbox"/> <u>g13383468</u> DC-SIGN [Homo sapiens]	532	e-150
<input checked="" type="checkbox"/> <u>g10179610</u> probable mannose-binding C-type lectin DC-SIGN [Hom	532	e-150

>g13383470 L-SIGN [Homo sapiens]
Length = 376

Score = 635 bits (1619), Expect = 0.0
Identities = 324/376 (86%), Positives = 325/376 (86%), Gaps = 51/376 (13%)

Query: 33 LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLQLKAAVGESEKSKLQEIYQELTQLK 92
LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLQLKAAVGESEKSKLQEIYQELTQLK
Sbjct: 61 LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLQLKAAVGESEKSKLQEIYQELTQLK 120

Query: 93 AAVGELPEKSKLQEIYQELTRLKAAVGEPEKSKLQEIYQELTRLKAAVGEPEKSKLQE 152
AAVGELPEKSKLQEIYQELTRLKAAVGEPEKSKLQEIYQELTRLKAAVGEPEKSKLQE
Sbjct: 121 AAVGELPEKSKLQEIYQELTRLKAAVGEPEKSKLQEIYQELTRLKAAVGEPEKSKLQE 180

Query: 153 IYQELT-----RLKAAVGEPLDQSKQQQIYQELTDLKTAFER 189
IYQELT +LKAAVGEPLDQSKQQQIYQELTDLKTAFER
Sbjct: 181 IYQELTELKAAVGEPEKSKLQEIYQELTQLKAAVGEPLDQSKQQQIYQELTDLKTAFER 240

Query: 190 LCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEVRAQLVVIKTAEQNFLLQLQTSRSN 249
LCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEVRAQLVVIKTAEQNFLLQLQTSRSN
Sbjct: 241 LCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEVRAQLVVIKTAEQNFLLQLQTSRSN 300

Query: 250 RFSWGLSDNLQEGTWQWVDGSPLSPSFQRYWNSGEPNNSGNEDCAEFGSGWNDNRCDV 309
RFSWGLSDNLQEGTWQWVDGSPLSPSFQRYWNSGEPNNSGNEDCAEFGSGWNDNRCDV
Sbjct: 301 RFSWGLSDNLQEGTWQWVDGSPLSPSFQRYWNSGEPNNSGNEDCAEFGSGWNDNRCDV 360

Query: 310 DNYWICKKPAACFRDE 325
DNYWICKKPAACFRDE
Sbjct: 361 DNYWICKKPAACFRDE 376

>g15383606 mDC-SIGN2 type I isoform [Homo sapiens]
Length = 399

Score = 621 bits (1585), Expect = e-177

Identities = 324/399 (81%), Positives = 325/399 (81%), Gaps = 74/399 (18%)

Query: 1 MSDSKEPRVQQLGLL-----GCLGHGALVLQLLSFML 32
MSDSKEPRVQQLGLL GCLGHGALVLQLLSFML
Sbjct: 1 MSDSKEPRVQQLGLLLEEDPTTSGIRLFPRDFQFQQIHGHKSSTGCLGHGALVLQLLSFML 60

Query: 33 LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLQLKAAVGESEKSKLQEIYQELTQLK 92
LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLQLKAAVGESEKSKLQEIYQELTQLK
Sbjct: 61 LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLQLKAAVGESEKSKLQEIYQELTQLK 120

Query: 93 AAVGELPEKSKL-----QEYQELTRLKAAVGEPEKSKLQE 129
AAVGELPEKSKL QEYQELTRLKAAVGEPEKSKLQE
Sbjct: 121 AAVGELPEKSKLQEYQELTRLKAAVGEPEKSKLQEYQELTRLKAAVGEPEKSKLQE 180

Query: 130 IYQELTR-----LKAAVGEPEKSKLQEYQELTRLKAAVGE 166
IYQELTR LKAAVGEPEKSKLQEYQELT+LKAAVGE
Sbjct: 181 IYQELTRLKAAVGEPEKSKLQEYQELTELKAAVGEPEKSKLQEYQELTQLKAAVGE 240

Query: 167 LPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEV 226
LPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEV
Sbjct: 241 LPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEV 300

Query: 227 AQLVVIKTAEQNFLLQLQTSRSNRFSWGLSDNLQEGTWQWVDGSPLSPSFQRYWNSGEP 286
AQLVVIKTAEQNFLLQLQTSRSNRFSWGLSDNLQEGTWQWVDGSPLSPSFQRYWNSGEP
Sbjct: 301 AQLVVIKTAEQNFLLQLQTSRSNRFSWGLSDNLQEGTWQWVDGSPLSPSFQRYWNSGEP 360

Query: 287 NNSGNEDCAEFGSGWNDNRCDVDNYWICKKPAACFFDE 325
NNSGNEDCAEFGSGWNDNRCDVDNYWICKKPAACFFDE
Sbjct: 361 NNSGNEDCAEFGSGWNDNRCDVDNYWICKKPAACFFDE 399

NCBI BLAST 2.2.26+ Linux x86_64 gcc4.6.20120914 [http://www.ncbi.nlm.nih.gov/blast/blast.cgi]

Score = 621 bits (1585), Expect = e-177

Identities = 324/399 (81%), Positives = 325/399 (81%), Gaps = 74/399 (18%)

Query: 1 MSDSKEPRVQQLGLL-----GCLGHGALVLQLLSFML 32
MSDSKEPRVQQLGLL GCLGHGALVLQLLSFML
Sbjct: 1 MSDSKEPRVQQLGLEEDPTTSGIRLFPRDFQFQQIHGHKSSTGCLGHGALVLQLLSFML 60

Query: 33 LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLKAAVGESEKSKLQEIYQELTQLK 92
LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLKAAVGESEKSKLQEIYQELTQLK
Sbjct: 61 LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLKAAVGESEKSKLQEIYQELTQLK 120

Query: 93 AAVGELPEKSKL-----QEIIYQELTRLKAAVGEPEKSKLQE 129
AAVGELPEKSKL QEIIYQELTRLKAAVGEPEKSKLQE
Sbjct: 121 AAVGELPEKSKLQEIIYQELTRLKAAVGEPEKSKLQEIIYQELTRLKAAVGEPEKSKLQE 180

Query: 130 IYQELTR-----LKAAVGEPEKSKLQEIIYQELTRLKAAVGE 166
IYQELTR LKAAVGEPEKSKLQEIIYQELT+LKAAVGE
Sbjct: 181 IYQELTRLKAAVGEPEKSKLQEIIYQELTELKAAVGEPEKSKLQEIIYQELTQLKAAVGE 240

Query: 167 LPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEV 226
LPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEV
Sbjct: 241 LPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEV 300

Query: 227 AQLVVIKTAAEQNFLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEP 286
AQLVVIKTAAEQNFLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEP
Sbjct: 301 AQLVVIKTAAEQNFLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEP 360

Query: 287 NNSGNEDCAEFGSGWNDNRCDVDNYWICKKPAACFRDE 325
NNSGNEDCAEFGSGWNDNRCDVDNYWICKKPAACFRDE
Sbjct: 361 NNSGNEDCAEFGSGWNDNRCDVDNYWICKKPAACFRDE 399

>g12084797 probable mannose binding C-type lectin DC-SIGNR [Homo
sapiens]
Length = 399

Score = 617 bits (1574), Expect = e-175

Identities = 323/399 (80%), Positives = 324/399 (80%), Gaps = 74/399 (18%)

Query: 1 MSDSKEPRVQQLGLL-----GCLGHGALVLQLLSFML 32
MSDSKEPRVQQLGLL GCLGHGALVLQLLSFML
Sbjct: 1 MSDSKEPRVQQLGLEEDPTTSGIRLFPRDFQFQQIHGHKSSTGCLGHGALVLQLLSFML 60

Query: 33 LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLKAAVGESEKSKLQEIYQELTQLK 92
LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLKAAVGESEKSKLQEIYQELTQLK
Sbjct: 61 LAGVLVAILVQVSKVPSSLSQEQSEQDAIYQNLTLKAAVGESEKSKLQEIYQELTQLK 120

Query: 93 AAVGELPEKSKL-----QEIIYQELTRLKAAVGEPEKSKLQE 129
AAVGELPEKSKL QEIIYQELTRLKAAVGEPEKSKLQE
Sbjct: 121 AAVGELPEKSKLQEIIYQELTRLKAAVGEPEKSKLQEIIYQELTRLKAAVGEPEKSKLQE 180

Query: 130 IYQELTF-----LFAAVGEPEKSKLQEIIYQELTRLKAAVGE 166
IYQELTF LFAAVGEPEKSKLQEIIYQELT+LFAAVGE
Sbjct: 181 IYQELTRLKAAVGEPEKSKLQEIIYQELTELKAAVGEPEKSKLQEIIYQELTQLKAAVGE 240

Query: 167 LPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEV 226
LPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEV
Sbjct: 241 LPDQSKQQQIYQELTDLKTAFERLCRHCPKDWTFQGNCFMSNSQRNWHDSVTACQEV 300

Query: 227 AQLVVIKTAAEQNFLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEP 286
AQLVVIKTAAEQNFLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEP
Sbjct: 301 AQLVVIKTAAEQNFLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEP 360

Sbjct: 361 NNSGNEDXAEFSGSGWNDNRCDVDNYWICKKPAACFRDE 399

>g15383614 sDC-SIGN2 type I isoform [Homo sapiens]
Length = 332

Score = 583 bits (1487), Expect = e-165

Identities = 293/332 (88%), Positives = 300/332 (90%), Gaps = 7/332 (2%)

Query: 1 MSDSKEPRVQQQLGLLG---CLGHGALVLQLLSFMLLAG----VLVAILVQVSKVPSSLSQ 53
MSDSKEPRVQQQLGLL L + F + G + +L VSKVPSSLSQ
Sbjct: 1 MSDSKEPRVQQQLGLEEDPTTSGIRLFPRDFQFQQIHGHKSSTVPFLLGPVSKVPSSLSQ 60

Query: 54 EQSEQDAIYQNLTLQKAAVGESEKSKLQEIYQELTLQKAAVGEPEKSKLQEIYQELTR 113
EQSEQDAIYQNLTLQKAAVGESEKSKLQEIYQELTLQKAAVGEPEKSKLQEIYQELTR
Sbjct: 61 EQSEQDAIYQNLTLQKAAVGESEKSKLQEIYQELTLQKAAVGEPEKSKLQEIYQELTR 120

Query: 114 LKAAVGEPEKSKLQEIYQELTRLKAAVGEPEKSKLQEIYQELTRLKAAVGEPLDQSKQ 173
LKAAVGEPEKSKLQEIYQELTRLKAAVGEPEKSKLQEIYQELT LKAAVGEPL++SK
Sbjct: 121 LKAAVGEPEKSKLQEIYQELTRLKAAVGEPEKSKLQEIYQELTELKAAVGEPEKSKL 180

Query: 174 QQIYQELTDLKTAFERLCRHCPCDWTFFQGNCFMSNSQRNWHDSVTACQEVRAQLVVIK 233
Q+IYQELT LK A ERLCRHCPCDWTFFQGNCFMSNSQRNWHDSVTACQEVRAQLVVIK
Sbjct: 181 QEIYQELTLKAAVERLCRHCPCDWTFFQGNCFMSNSQRNWHDSVTACQEVRAQLVVIK 240

Query: 234 TAEQNFLLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEPNNSGNED 293
TAEQNFLLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEPNNSGNED
Sbjct: 241 TAEQNFLLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEPNNSGNED 300

Query: 294 CAEFSGSGWNDNRCDVDNYWICKKPAACFRDE 325
CAEFSGSGWNDNRCDVDNYWICKKPAACFRDE
Sbjct: 301 CAEFSGSGWNDNRCDVDNYWICKKPAACFRDE 332

>g8572543 membrane-associated lectin type-C [Homo sapiens]
Length = 404

Score = 532 bits (1356), Expect = e-150

Identities = 281/392 (71%), Positives = 301/392 (76%), Gaps = 71/392 (18%)

Query: 1 MSDSKEPRVQQQLGLL-----GCLGHGALVLQLLSFMLLAGVLVAI 40
MSDSKEPR+QQLGLL GCLGHG LVLQLLSF LLAG +
Sbjct: 1 MSDSKEPRLLQQLGLEEEQLRGLGFRQTRGYKSLAGCLGHGPLVLQLLSFTLLAG----L 56

Query: 41 LVQVSKVPSSLSQEQSEQDAIYQNLTLQKAAVGESEKSKLQEIYQELTLQKAAVGEPE 100
LVQVSKVPSS+SQEQS QDAIYQNLTLQKAAVGESEKSKLQEIYQELTLQKAAVGEPE
Sbjct: 57 LVQVSKVPSSISQEQSRQDAIYQNLTLQKAAVGESEKSKLQEIYQELTLQKAAVGEPE 116

Query: 101 KSKL-----QEYIYQELTRL 114
KSKL QEYIYQELTRL
Sbjct: 117 KSKLQEYIYQELTRLKAAVGEPEKSKLQEYIYQELTWLKAAVGEPEKSKMQEYIYQELTRL 176

Query: 115 KAAVGEPEKSKLQEYIYQELTRLKAAVGEPEKSKLQEYIYQELTRLKAAVGEPLDQSKQQ 174
KAAVGEPEKSK QEYIYQELTRLKAAVGEPEKSK QEYIYQELTRLKAAVGEPL++SKQQ
Sbjct: 177 KAAVGEPEKSKQEYIYQELTRLKAAVGEPEKSKQEYIYQELTRLKAAVGEPEKSKQQ 236

Query: 175 QIYQELTDLKTAFERLCRHCPCDWTFFQGNCFMSNSQRNWHDSVTACQEVRAQLVVIKT 234
+IYQELT LK A ERLC CP +WTFFQGNCFMSNSQRNWHDS+TAC+EV AQLVVIK+
Sbjct: 237 RIYQELTLQKAAVERLCHPCPEWTFQGNCFMSNSQRNWHDSITACKEVGAQLVVIKS 296

Query: 235 ABEQNFLLQLQTSRSNRFSSWGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEPNNSGNEDC 294
ABEQNFLLQLQ+SRNRF+WMGLSDLNQEGTWQWVDGSPSPSFQRYWNSGEPNNSGNEDC

AEFSG+GWND++C++ +WICKK AA C RDE
Sbjct: 357 AEFSGNGWDDKCNLAKFWICKKSAASCSRDE 388

>g17049084 unnamed protein product [Homo sapiens]
Length = 404

Score = 532 bits (1356), Expect = e-150
Identities = 281/392 (71%), Positives = 301/392 (76%), Gaps = 71/392 (18%)

Query: 1 MSDSKEPRVQQLGLL-----GCLGHGALVLQLLSFMLLAGVLVAI 40
MSDSKEPR+QQLGLL GCLGHG LVLQLLSF LLAG +
Sbjct: 1 MSDSKEPRLQQLGLLEEEQLRGLGFRQTRGYKSLAGCLGHGPLVLQLLSFTLLAG---L 56

Query: 41 LVQVSKVPSSLSQEQSEQDAIYQNLTLKAAVGESEKSKLQEIYQELTLKAAVGEPE 100
LVQVSKVPSS+SQEQS QDAIYQNLTLKAAVGESEKSKLQEIYQELTLKAAVGEPE
Sbjct: 57 LVQVSKVPSSISQEQSRQDAIYQNLTLKAAVGESEKSKLQEIYQELTLKAAVGEPE 116

Query: 101 KSKL-----QEIIYQELTRL 114
KSKL QEIIYQELTRL
Sbjct: 117 KSKLQEIIYQELTRLKAAVGEPEKSKLQEIYQELTWLKAAVGEPEKSKMQEIYQELTRL 176

Query: 115 KAAVGEPEKSKLQEIYQELTRLKAAVGEPEKSKLQEIYQELTRLKAAVGEPEPDQSKQQ 174
KAAVGEPEKSK QEIIYQELTRLKAAVGEPEKSK QEIIYQELTRLKAAVGEPE++SKQQ
Sbjct: 177 KAAVGEPEKSKQQEIYQELTRLKAAVGEPEKSKQQEIYQELTRLKAAVGEPEKSKQQ 236

Query: 175 QIIYQELTDLKTAFERLCRHCPRKDWTFQNGCYFMSNSQRNWHDSVTACQEVRAQLVVIKT 234
+IYQELT LK A ERLC CP +WTFQNGCYFMSNSQRNWHDS+TAC+EV AQLVVIK+
Sbjct: 237 EIIYQELTLKAAVERLCHPCPWEWTFQNGCYFMSNSQRNWHDSITACKEVGAQLVVIKS 296

Query: 235 AEEQNFLQLQTSRSNRFSWMGLSDLNQEGTWQWVDGSPLSPSFQRYWNSGEPNNNGNEDC 294
AEEQNFLQLQ+SRSNRF+WMGLSDLNQEGTWQWVDGSPL PSF++YWN GEPNN G EDC
Sbjct: 297 AEEQNFLQLQSSRSNRFTWMGLSDLNQEGTWQWVDGSPLLPSPFKQYWNRGEPNNVGEEDC 356

Query: 295 AEFSGSGWNDNRCDVDNYWICKKPAA-CFRDE 325
AEFSG+GWND++C++ +WICKK AA C RDE
Sbjct: 357 AEFSGNGWDDKCNLAKFWICKKSAASCSRDE 388

>g15281073 mDC-SIGN1A type I isoform [Homo sapiens]
Length = 404

Score = 532 bits (1356), Expect = e-150
Identities = 281/392 (71%), Positives = 301/392 (76%), Gaps = 71/392 (18%)

Query: 1 MSDSKEPRVQQLGLL-----GCLGHGALVLQLLSFMLLAGVLVAI 40
MSDSKEPR+QQLGLL GCLGHG LVLQLLSF LLAG +
Sbjct: 1 MSDSKEPRLQQLGLLEEEQLRGLGFRQTRGYKSLAGCLGHGPLVLQLLSFTLLAG---L 56

Query: 41 LVQVSKVPSSLSQEQSEQDAIYQNLTLKAAVGESEKSKLQEIYQELTLKAAVGEPE 100
LVQVSKVPSS+SQEQS QDAIYQNLTLKAAVGESEKSKLQEIYQELTLKAAVGEPE
Sbjct: 57 LVQVSKVPSSISQEQSRQDAIYQNLTLKAAVGESEKSKLQEIYQELTLKAAVGEPE 116

Query: 101 KSKL-----QEIIYQELTRL 114
KSKL QEIIYQELTRL
Sbjct: 117 KSKLQEIIYQELTRLKAAVGEPEKSKLQEIYQELTWLKAAVGEPEKSKMQEIYQELTRL 176

Query: 115 KAAVGEPEKSKLQEIYQELTRLKAAVGEPEKSKLQEIYQELTRLKAAVGEPEPDQSKQQ 174
KAAVGEPEKSK QEIIYQELTRLKAAVGEPEKSK QEIIYQELTRLKAAVGEPE++SKQQ
Sbjct: 177 KAAVGEPEKSKQQEIYQELTRLKAAVGEPEKSKQQEIYQELTRLKAAVGEPEKSKQQ 236

Query: 235 AEEQNFLQLQTSRSNRF+WMGLSDLNQEGLTWQWVDGSPSPSFQRYWNSGEPNNNSGNEDC 294
AEEQNFLQLQ+SRSNRF+WMGLSDLNQEGLTWQWVDGSPSPSF++YWN GEPNN G EDC
Sbjct: 297 AEEQNFLQLQSSRSNRF+WMGLSDLNQEGLTWQWVDGSPSPSFQRYWNRGEPNNVGEEDC 356

Query: 295 AEFSGSGWNDNRCDVDNYWICKKPAA-CFRDE 325
AEFSG+GWND++C++ +WICKK AA C RDE
Sbjct: 357 AEFSGNGWDDKCNLAKFWICKKSAASCSRDE 388

>g13383468 DC-SIGN [Homo sapiens]
Length = 404

Score = 532 bits (1356), Expect = e-150
Identities = 281/392 (71%), Positives = 301/392 (76%), Gaps = 71/392 (18%)

Query: 1 MSDSKEPRVQQLGLL-----GCLGHGALVLQLLSFMLLAGVLVAI 40
MSDSKEPR+QQLGLL GCLGHG LVLQLLSF LLAG +
Sbjct: 1 MSDSKEPRLQQLGLEEEQLRGLGFRQTRGYKSLAGCLGHGPLVLQLLSFTLLAG----L 56

Query: 41 LVQVSKVPSSLSQEQSEQDAIYQNLTQLKAAVGELSEKSKLQEIYQELTQLKAAVGELPE 100
LVQVSKVPSS+SQEQS QDAIYQNLTQLKAAVGELSEKSKLQEIYQELTQLKAAVGELPE
Sbjct: 57 LVQVSKVPSSISQEQSRQDAIYQNLTQLKAAVGELSEKSKLQEIYQELTQLKAAVGELPE 116

Query: 101 KSKL-----QEIYQELTRL 114
KSKL QEIYQELTRL
Sbjct: 117 KSKLQEIYQELTRLKAAVGELPEKSKLQEIYQELTWLKAAVGELPEKSKMQEIYQELTRL 176

Query: 115 KAAVGELPEKSKLQEIYQELTRLKAAVGELPEKSKLQEIYQELTRLKAAVGELPDQSKQQ 174
KAAVGELPEKSK QEIYQELTRLKAAVGELPEKSK QEIYQELTRLKAAVGELP++SKQQ
Sbjct: 177 KAAVGELPEKSKQQEIYQELTRLKAAVGELPEKSKQQEIYQELTRLKAAVGELPEKSKQQ 236

Query: 175 QIYQELTDLKTAFERLCRHCPKDWTFQNGCYFMSNSQRNWHDSVTACQEVRAQLVVIKT 234
+IYQELT LK A ERLC CP +WTFQNGCYFMSNSQRNWHDS+TAC+EV AQLVVIK+
Sbjct: 237 EIIYQELTQLKAAVERLCHPCPWEWTFQNGCYFMSNSQRNWHDSITACKEVGAQLVVIKS 296

Query: 235 AEEQNFLQLQTSRSNRF+WMGLSDLNQEGLTWQWVDGSPSPSFQRYWNSGEPNNNSGNEDC 294
AEEQNFLQLQ+SRSNRF+WMGLSDLNQEGLTWQWVDGSPSPSF++YWN GEPNN G EDC
Sbjct: 297 AEEQNFLQLQSSRSNRF+WMGLSDLNQEGLTWQWVDGSPSPSFQRYWNRGEPNNVGEEDC 356

Query: 295 AEFSGSGWNDNRCDVDNYWICKKPAA-CFRDE 325
AEFSG+GWND++C++ +WICKK AA C RDE
Sbjct: 357 AEFSGNGWDDKCNLAKFWICKKSAASCSRDE 388

>g10179610 probable mannose-binding C-type lectin DC-SIGN [Homo sapiens]
Length = 404

Score = 532 bits (1356), Expect = e-150
Identities = 281/392 (71%), Positives = 301/392 (76%), Gaps = 71/392 (18%)

Query: 1 MSDSKEPRVQQLGLL-----GCLGHGALVLQLLSFMLLAGVLVAI 40
MSDSKEPR+QQLGLL GCLGHG LVLQLLSF LLAG +
Sbjct: 1 MSDSKEPRLQQLGLEEEQLRGLGFRQTRGYKSLAGCLGHGPLVLQLLSFTLLAG----L 56

Query: 41 LVQVSKVPSSLSQEQSEQDAIYQNLTQLKAAVGELSEKSKLQEIYQELTQLKAAVGELPE 100
LVQVSKVPSS+SQEQS QDAIYQNLTQLKAAVGELSEKSKLQEIYQELTQLKAAVGELPE
Sbjct: 57 LVQVSKVPSSISQEQSRQDAIYQNLTQLKAAVGELSEKSKLQEIYQELTQLKAAVGELPE 116

Query: 101 KSKL-----QEIYQELTRL 114
KSKL QEIYQELTRL
Sbjct: 117 KSKLQEIYQELTRLKAAVGELPEKSKLQEIYQELTWLKAAVGELPEKSKMQEIYQELTRL 176

Sbjct: 177 KAAVGELPEKSKQQEIYQELTRLKAAVGELPEKSKQQEIYQELTRLKAAVGELPEKSKQQ 236

Query: 175 QIYQELTDLKTAFERLCRHCPKDWTFFQGNCYFMSNSQRNWHDSVTACQEVRAQLVVIKT 234
+IYQELT LK A ERLC CP +WTFFQGNCYFMSNSQRNWHDS+TAC+EV AQLVVIK+

Sbjct: 237 EIYQELTQLKAAVERLCHPCPWEWTFQGNCYFMSNSQRNWHDSITACKEVGAQLVVIKS 296

Query: 235 AEEQNFLQLQTSRNRFSWMGLSDLNQEGTWQWVDGSPLSPSFQRYWNSGEPNNSGNEDC 294
AEEQNFLQLQ+SRSNRF+WMGLSDLNQEGTWQWVDGSPL PSF++YWN GEPNN G EDC

Sbjct: 297 AEEQNFLQLQSSRSNRFTWMGLSDLNQEGTWQWVDGSPLPSFKQYWNRGEPNNVGEDC 356

Query: 295 AEFGSGGWNDNRCDVDNYWICKKPAA-CFRDE 325
AEFGSG+GWND++C++ +WICKK AA C RDE

Sbjct: 357 AEFGSGGWNDKCNLAKFWICKKSAASCSRDE 388

Database: genpept131

Posted date: Sep 9, 2002 12:31 PM

Number of letters in database: 348,344,575

Number of sequences in database: 1,135,942

Lambda	K	H
0.315	0.130	0.384

Gapped

Lambda	K	H
0.270	0.0470	0.230

Matrix: BLOSUM62

Gap Penalties: Existence: 11, Extension: 1

Number of Hits to DB: 213284592

Number of Sequences: 1135942

Number of extensions: 8629475

Number of successful extensions: 36326

Number of sequences better than 10.0: 1471

Number of HSP's better than 10.0 without gapping: 638

Number of HSP's successfully gapped in prelim test: 888

Number of HSP's that attempted gapping in prelim test: 32786

Number of HSP's gapped (non-prelim): 2270

length of query: 325

length of database: 348,344,575

effective HSP length: 59

effective length of query: 266

effective length of database: 281,323,997

effective search space: 74832183202

effective search space used: 74832183202

T: 11

A: 40

X1: 16 (7.3 bits)

X2: 38 (14.8 bits)

X3: 64 (24.9 bits)

S1: 41 (21.5 bits)

Graphical Viewer...

Submit sequences to: BLAST2



Submit

Reset



hmmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /data/isb2k/blastdb/Pfam72/Pfam72

Sequence file: /u/legal/jennyb/pf636.seq

Query: 3344986CD1

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
lectin_c	Lectin C-type domain	139.5	5.9e-38	1
Ribosomal_L29	Ribosomal L29 protein	-15.3	9.1	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
Ribosomal_L29	1/1	85	152	1	64	-15.3	9.1
lectin_c	1/1	211	317	1	125	139.5	5.9e-38

Alignments of top-scoring domains:

Ribosomal_L29: domain 1 of 1, from 85 to 152: score -15.3, E = 9.1

```

      *->akELRelsde..EL..eeleelKrELfeLRAfqaAtGqLenPhrIk
      ++EL +l+ + +EL+++ +l e+ +EL L+  aA+G+L  +++
3344986CD1    85    YQELTQLKAAvgELpeKSKLQEIIYQELTRLK---AAVGELPEKSKLQ 128

      evRkrIARilTv...lnErklxae<-*
      e+ +++ R++ + ++l E+ + +e
3344986CD1    129  EIIYQELTRLKAAvgELPEKSKLQE    152

```

lectin_c: domain 1 of 1, from 211 to 317: score 139.5, E = 5.9e-38

```

      *->esktWaeAelaCqkegghAHLvsIqsaeEqsfvvaflltsltkksnty
      ++++W+++ +aCq+ ++ Lv+I aeEq  +fl+ t++sn
3344986CD1    211    SQRNWHDSVTACQEVRAQ--LVVIKTAEEQ----NFLQLQTSRSNRF 251

      aWIGLtdintegtwwegwetdgspvnyt..enWapgePnnrghGgnEd
      W+GL+d n+egtw+w  +dgsp++ + +++W++gePnn gn  Ed
3344986CD1    252    SWMGLSDLNQEGTWQW---VDGSPLSPsfqRYWNSGEPNNSGN---ED 293

      CveiytdtdflaGkwnDepCdsklpvyvCef<-*
      C+e++++ WnD+ Cd+ + ++C++
3344986CD1    294    CAEFSGS-----GWNDNRCDVDNYWICKK    317

```

BLIMPS (BBlocks IMProved Searcher) Version 3.4 2000/06
(C) Copyright 1993-2000, Fred Hutchinson Cancer Research Center

Probe Sequence: 3344986CD1

Probe Size: 325 Amino Acids

Probe File: /u/legal/jennyb/WIP.23834/3344986CD1.fasta

Target File (s) : /opt/pat/prod/SOL/blimps-3.4/dat/blocks.dat

Records Searched: 4071

Scores Done: 4071

Alignments Done: 1439864

AC#	Description	Strength	Score
RF AA#			
IPB001304A	C-type lectin domain	1072	1164
0 197	WTFFQGNCYFMSNSQRNWHDSVTAC		
IPB001304B	C-type lectin domain	1000	1163
0 252	WMGLSDLNQEGTW		

MOTIFS from: /u/legal/jennyb/WIP.23723/3344986CD1.gcg

Mismatches: 0 January 30, 2003 14:55 ..

3344986CD1.gcg Check: 7873 Length: 325 ! >3344986CD1

C _ T y p e _ L e c t i n _ 1

C(L,I,V,M,F,Y,A,T,G)x{5,12}(W,L)x(D,N,S,R)x2Cx{5,6}(F,Y,W,L,I,V,S,T,A)(L,I,V,M,
,S,T,A)C

C(A)x{6}(W)x(D)x{2}Cx{5}(W)(I)C

294: SGNE D CAEFSGSGWNDNRCDVDNYWIC KKPAA

[illegible]

S	G	N	E	D	C	A	E	F	S	G	S	G	W	N	D	N	R	C	D	V	D	N	Y	W	I	C	K	SEQIDNO_12
V	G	E	E	D	C	A	E	F	S	G	N	G	W	N	D	D	K	C	N	L	A	K	F	W	I	C	K	GENESEQ_AAR32188
V	G	E	E	D	C	A	E	F	S	G	N	G	W	N	D	D	K	C	N	L	A	K	F	W	I	C	K	g8572543
S	G	N	E	D	C	A	E	F	S	G	S	G	W	N	D	N	R	C	D	V	D	N	Y	W	I	C	K	g13383470

A A - C F R D E E
 A A S C S R D E E
 A A S C S R D E E
 A A - C F R D E E

Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes

(Human Genome Project/DNA chip/gene discovery/T cell)

MARK SCHENA*†, DARI SHALON‡, RENU HELLER*, ANDREW CHAI*, PATRICK O. BROWN§, AND RONALD W. DAVIS*

*Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305; †Syntex, Palo Alto, CA 94306; and ‡Department of Biochemistry and Howard Hughes Medical Institute, Beckman Center, Stanford University Medical Center, Stanford, CA 94305

Contributed by Ronald W. Davis, June 26, 1996

ABSTRACT Microarrays containing 1046 human cDNAs of unknown sequence were printed on glass with high-speed robotics. These 1.0-cm² DNA "chips" were used to quantitatively monitor differential expression of the cognate human genes using a highly sensitive two-color hybridization assay. Array elements that displayed differential expression patterns under given experimental conditions were characterized by sequencing. The identification of known and novel heat shock and phorbol ester-regulated genes in human T cells demonstrates the sensitivity of the assay. Parallel gene analysis with microarrays provides a rapid and efficient method for large-scale human gene discovery.

Biology has entered the genome era (1). Complete genome sequences for all of the model organisms and human will probably be available by the year 2003 (2). Torrents of human expressed sequence tags (ESTs) provide a starting point for elucidating the function of tens of thousands of cognate genes (3). Genome analysis will provide insights into growth, development, differentiation, homeostasis, aging, and the onset of diseases (1–3). A detailed understanding of the human genome will require the implementation of sophisticated methods for gene expression analysis and gene discovery.

Recently, a microarray-based method for high-throughput monitoring of plant gene expression was described (4). This "chip"-based approach involved using microarrays of cDNA clones as gene-specific hybridization targets to quantitatively measure expression of the corresponding plant genes (4, 5). A two-color fluorescence labeling and detection scheme facilitated sensitive differential expression analysis of different plant tissues (4, 5). The efficiency of this approach for studies in higher plants suggested the use of this method for human genome analysis (4-7). Here, we report the use of cDNA microarrays for human gene expression monitoring, biological investigation, and gene discovery.

MATERIALS AND METHODS

Human cDNA Clones. The cDNA library was made with mRNA from human peripheral blood lymphocytes transformed with the Epstein-Barr virus. Inserts >600 bp were cloned into the lambda vector λ YES-R to generate 10^7 – 10^8 recombinants. Bacterial transformants were obtained by infecting *E. coli* strain JM107/ λ KC. Colonies were picked at random and propagated in a 96-well format, and miniprep DNA was prepared by alkaline lysis using REAL preps (Qiagen, Chatsworth, CA). Inserts were amplified by PCR in a 96-well format using primers (PAN132, 5'-CCTC-TATACTTTAACGTCAAGG; and PAN133, 5'-TTGTGTG-GAATTGTGAGCGG) complementary to the λ YES

(Glen Research, Sterling, VA) on the 5' end. PCR products were purified in a 96-well format using QIAquick columns (Qiagen).

Microarray Preparation. Amino-modified PCR products were suspended at a concentration of 0.5 ng/ml in 3× standard saline citrate (SSC) and arrayed from 96-well microtiter plates onto silylated microscope slides (CEL Associates, Houston) using high-speed robotics (4–7). A total of 1056 cDNAs, representing 1046 human clones and 10 *Arabidopsis* controls, were arrayed in 1.0-cm² areas. Printed arrays were incubated for 4 hr in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in H₂O for 1 min, and once for 5 min in sodium borohydride solution (1.0 g of NaBH₄ dissolved in 300 ml of PBS and 100 ml of 100% ethanol). The arrays were submerged in H₂O for 2 min at 95°C, transferred quickly into 0.2% SDS for 1 min, rinsed twice in H₂O, air dried, and stored in the dark at 25°C.

Fluorescent Probes. Tissue mRNAs were purchased (CLONTECH). Jurkat mRNA was isolated as described by Schena *et al.* (4). Probes were made as described (4) with several modifications. The reverse transcriptase used here was Superscript II RNase H- (GIBCO). The Cy5-dCTP was purchased from Amersham. Each reverse transcription reaction contained 3.0 μg of total human mRNA. *Arabidopsis* control mRNAs were made by *in vitro* transcription of cloned HAT4, HAT22, and YesAt-23 cDNAs (4, 8, 9) using an RNA Transcription Kit (Stratagene). For quantitation, the mRNAs were doped into the reverse transcription reaction at ratios of 1:100,000, 1:10,000, and 1:1,000 (wt/wt) respectively. Following the reverse transcription step, samples were treated with 2.5 μl of 1 M sodium hydroxide for 10 min at 37°C, then neutralized by adding 2.5 μl of 1 M Tris-HCl (pH 6.8) and 2.0 μl of 1 M HCl. Probe mixtures contained cDNA products derived from 3 μg of total mRNA, suspended in 5.0 μl of hybridization buffer (5 \times SSC plus 0.2% SDS).

Hybridization and Scanning. Probes were hybridized to 1.0-cm² microarrays under a 14 × 14 mm glass coverslip for 6–12 hr at 60°C in a custom-built hybridization chamber (4–7). Arrays were washed for 5 min at room temperature (25°C) in low stringency wash buffer (1 × SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 × SSC/0.2% SDS). Arrays were scanned in 0.1 × SSC using a fluorescence laser scanning device (4–7), fitted with a custom filter set (Chroma Technology, Brattleboro, VT). Accurate differential expression measurements (i.e., final fluorescence ratios) were obtained by taking the average of the ratios of two independent hybridizations.

14-00000. This article was destroyed in part by back issue
 deletion. This article must therefore be hereby marked "adversely" in
 accordance with 38 U.S.C. § 1734 solely to indicate this fact.

Existed in the GenBank data base (accession nos. U06604-U06660).
In whom reprint requests should be addressed: Vladimir Schevsky,
cmgm@stanford.edu.

Cell Culture. Jurkat cells were grown in a tissue culture incubator (37°C and 5% CO₂) in RPMI medium supplemented with 10% fetal bovine serum, 100 µg of streptomycin per ml, and 500 units of penicillin per ml. Heat shock corresponded to a 4-hr incubation at 43°C. Phorbol ester treated cells were grown for 4 hr in the presence of 50 ng of phorbol 12-myristate 13-acetate (PMA) per ml.

RNA Blotting. Dot blots were performed as described (4).

DNA Sequencing. Sequences were obtained using the PAN132 and PAN133 primers and a 373A automated sequencer, according to the instructions of the manufacturer (Applied Biosystems).

Computer Graphics and Informatics. Pseudocolor representations of fluorescent images were made with National Institutes of Health IMAGE software (version 1.52). Software for differential expression representations was purchased from Imaging Research (St. Catherine's, ON, Canada). Sequence searches were made to the nonredundant nucleotide data base at the National Center for Biotechnology Information (NCBI) using Macintosh BLAST software. The EST data base was accessed via the World Wide Web (<http://www.ncbi.nlm.nih.gov/>).

RESULTS

Gene Discovery and the Heat Shock Response. Microarrays were used to examine the heat shock response in cultured human T (Jurkat) cells. Control (37°C) and heat-treated (43°C) cells were harvested and lysed, and total mRNA from the two cell samples was labeled by reverse transcriptase incorporation of fluorescein- and Cy5-dCTP, respectively. In a second set of labeling reactions, the fluorescent groups were "swapped" such that samples from control and heat-treated

samples were labeled with Cy5- and fluorescein-dCTP, respectively. Each pair of fluorescent probes was hybridized to a 1056-element microarray. The arrays were washed at high stringency and scanned with a confocal laser scanning device to detect emission of the two fluorescent groups.

Hybridization signals were observed to >95% of the human cDNA array elements, but not to any of the *Arabidopsis* negative controls (Fig. 1). Fluorescence intensities spanned more than three orders of magnitude for the 1046 array elements surveyed (Fig. 1). Comparative expression analysis of heat shocked versus control cells in the two experiments revealed 17 array elements that displayed altered fluorescence ratios of ≥ 2.0 -fold (Figs. 1 and 2A). Of the 17 putative differentially expressed genes, 11 were induced by heat shock treatment and 6 displayed modest repression (Figs. 1 and 2A).

To determine the identity of the heat-regulated genes, cDNAs corresponding to each of the 17 array elements were sequenced on the proximal and distal end. Data base searches revealed perfect matches for 14 of the 17 clones, and in each case proximal and distal cDNA sequences mapped to the same gene (Table 1). Of the 1046 human genes examined on the microarray, the five most highly induced in heat-treated cells were heat shock protein 90 α (hsp90 α), dnaJ, hsp90 β , polyubiquitin, and t-complex polypeptide-1 (tcp-1) (Table 1). Three of the 17 clones did not match any entry in the public data base, though one of the clones (B7) exhibited significant homology to an EST from *Caenorhabditis elegans* (Table 1). Each of the novel sequences (B7-B9) exhibited ~ 2 -fold induction (Table 1) and relatively low-level expression (Table 2).

To confirm the microarray results, mRNA levels for each of the genes were measured by RNA blotting. Each of the genes that displayed heat shock induction, including the three novel

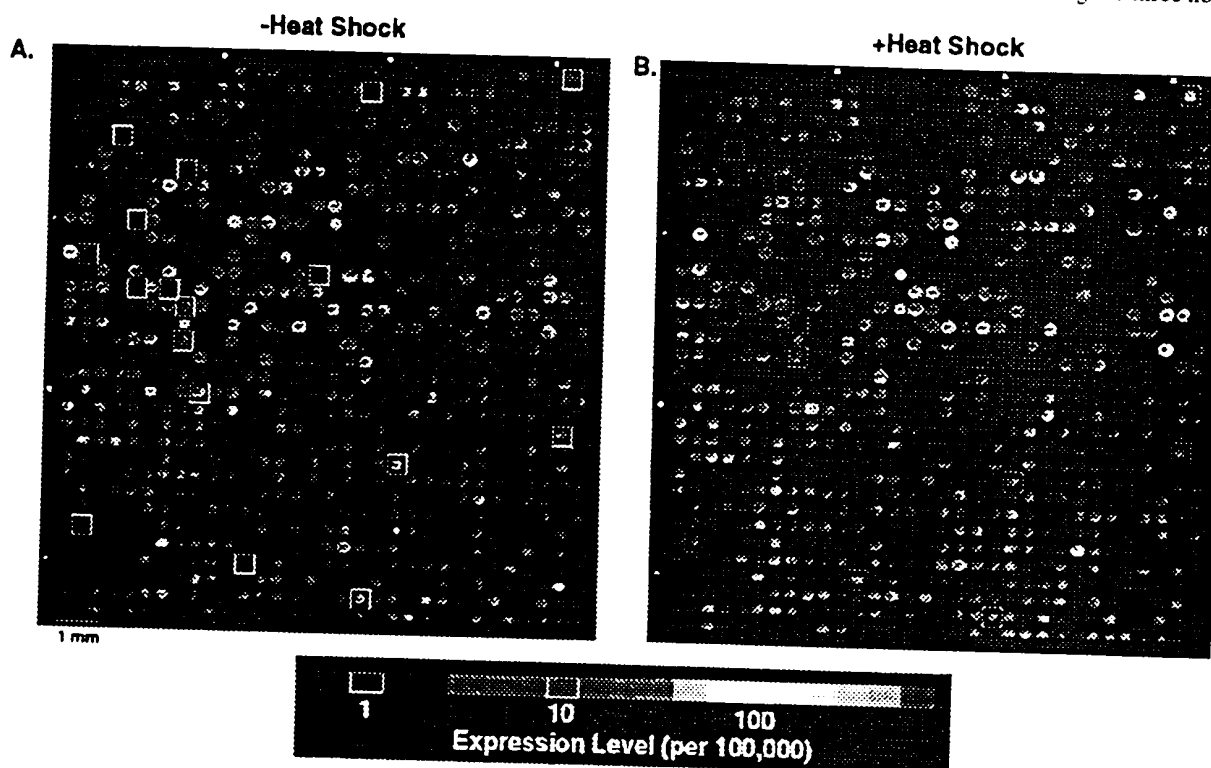


FIG. 1. Human gene expression monitored on a microarray. Fluorescent scans represented in a pseudocolor scale. (A) Control scan of Jurkat cells. (B) Heat shock scan. The color bar indicates the expression level (per 100,000) of the genes. The color bar was calibrated in separate experiments using known quantities (wt/wt) of *Arabidopsis* control mRNAs added to the labeling reaction. Microarray rows and columns (at the top) are demarcated at 10 element increments (white circles). (Bar = 1 mm.)

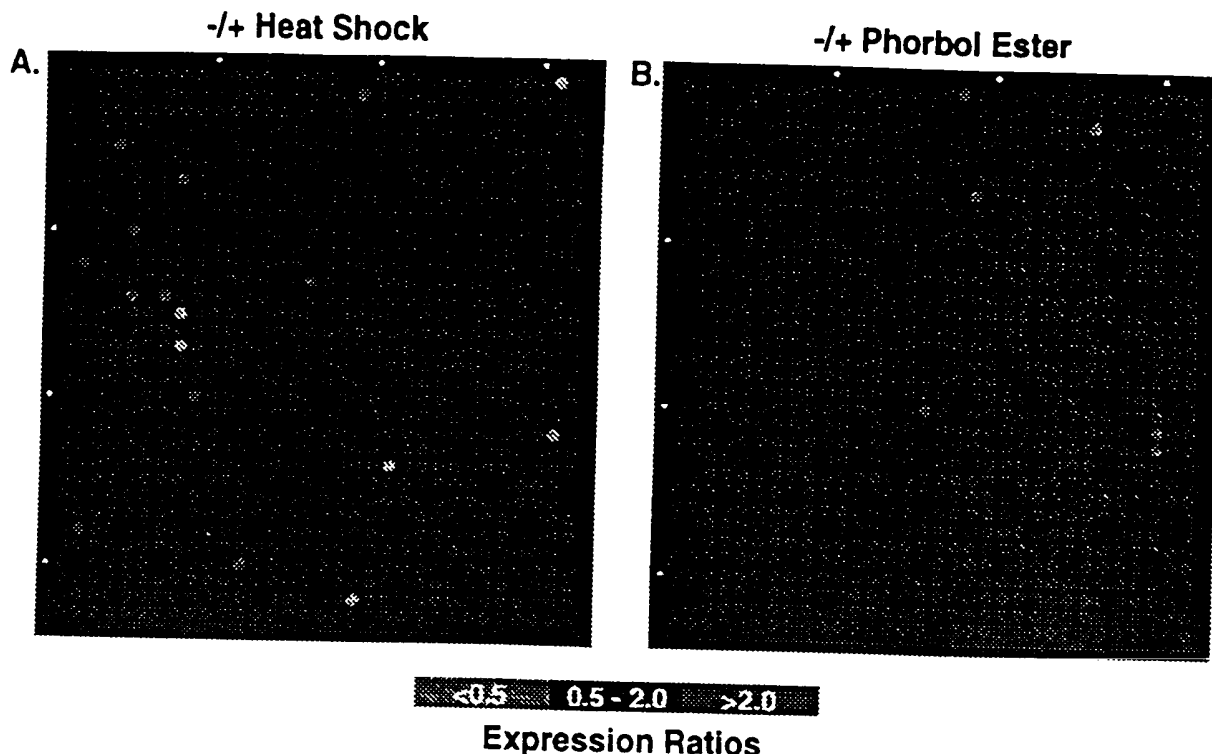


FIG. 2. Elemental displays of activated and repressed genes. Fluorescence ratios of two-color microarray scans (Fig. 1) are depicted schematically. Fluorescein-labeled probes from Jurkat cells subjected to (A) heat shock or (B) phorbol ester treatment were compared with Cy5-labeled probes from untreated cells. In a second set of reactions, the fluorescent groups were swapped (see text). The data represent the average of the ratios from two hybridizations, excluding values in which the difference of the two ratios was greater than half the average ratio. The color bar corresponds to expression ratios, which are independent of the absolute expression level of a given gene.

Table 1. Microarray elements corresponding to differentially expressed genes

Clone	Row	Column	Ratio	Blast identity	Accession no.
B1	24	21	0.5	CYC oxidase III	J01415, J01415
B2	1	31	0.5	β -Actin	NR, X00351
B3	15	8	0.5	CYC oxidase III	J01415, J01415
B4	32	19	0.5	CYC oxidase III	J01415, J01415
B5	17	8	0.5	CYC oxidase III	J01415, J01415
B6	22	31	0.5	β -Actin	NR, X00351
B7*	5	4	2.0	Novel†	U56653, U56654
B8	2	19	2.0	Novel†	U56655, U56656
B9	14	5	2.2	Novel†	U56657, U56658
B10	7	8	2.4	Polyubiquitin	X04803, X04803
B11	12	2	2.4	TCP-1	X52882, X52882
B12	28	2	2.5	Polyubiquitin	M17597, M17597
B13	14	7	2.5	Polyubiquitin	X04803, X04803
B14	20	9	2.6	HSP90 β	M16660, M16660
B15	30	12	4.0	DnaJ homolog	D13388, D13388
B16	10	5	5.8	HSP90 α	X07270, X07270
B17	13	16	6.3	HSP90 α	M27024, X15183
B18	7	19	2.0	β 2-microglobulin	S54761, M30683
B19	21	30	2.1	Novel†	U56659, U56660
B20	3	26	2.2	β 2-microglobulin	S54761, M30683
B21	1	18	2.6	PGK	M11968, L00160
B22	22	30	3.5	NF- κ B1	Z47744, M55643
B23	20	16	19	PAC-1	L11329, L11329

Clone name, array position (Fig. 1), fluorescence ratio, sequence identity, and accession number of cDNAs that manifested a differential expression pattern with probes prepared from heat shock- (B1-17) or phorbol ester-treated (B18-23) Jurkat cells. Clones showing >98% identity over 300 nucleotides were assumed to be identical to known sequences. All genes are nuclear except CYC oxidase III (mitochondrial). Accession numbers of cDNA sequences are from GenBank.

* B7 is identical to an EST from *S. cerevisiae* (U56653).
† No match in the public data bases.

Table 2. Human gene expression monitored by microarray and RNA blot analyses

Clone	Blast identity	Expression level, per 10 ⁵ mRNAs			
		Microarray	Ratio	RNA blot	Ratio
B1	CYC oxidase III	92/46	0.5	100/80	0.8
B2	β -Actin	240/120	0.5	270/280	1.0
B3	CYC oxidase III	36/18	0.5	ND	ND
B4	CYC oxidase III	76/38	0.5	ND	ND
B5	CYC oxidase III	62/31	0.5	ND	ND
B6	β -Actin	180/89	0.5	ND	ND
B7	Novel (weakly to D76026)	1.3/2.6	2.0	0.77/1.8	2.3
B8	Novel	2.0/4.0	2.0	1.5/3.4	2.3
B9	Novel	0.8/1.8	2.2	1.2/1.8	1.5
B10	Polyubiquitin	0.8/1.9	2.4	25/89	3.6
B11	TCP-1	2.3/5.5	2.4	7.1/27	3.8
B12	Polyubiquitin	0.8/2.0	2.5	ND	ND
B13	Polyubiquitin	1.7/4.3	2.5	ND	ND
B14	HSP90 β	75/200	2.6	30/120	4.0
B15	DnaJ homolog	1.0/4.0	4.0	1.6/13	8.1
B16	HSP90 α	0.6/3.5	5.8	3.2/29	9.1
B17	HSP90 α	0.8/5.0	6.3	8.6/62	7.2
B18	β_2 -microglobulin	1.0/2.0	2.0	5.4/15	2.8
B19	Novel	1.2/2.5	2.1	4.5/9.5	2.5
B20	β_2 -microglobulin	2.7/5.9	2.2	ND	ND
B21	Phosphoglycerate kinase	2.4/6.2	2.6	4.7/9.2	2.0
B22	NF-KB1	1.7/6.0	3.5	0.65/4.7	7.2
B23	PAC-1	0.5/9.5	19	0.21/15	71

Shown are expression levels per 100,000 mRNAs (wt/wt) of genes assayed with a microarray (Fig. 1) or RNA blot. Ratios correspond to values from cells subjected to heat shock (B1-17) or phorbol ester treatment (B18-23) relative to untreated cells. Clone and gene names are given in Table 1. ND, not determined.

sequences, exhibited elevated mRNA levels by dot blot analysis (Table 2). In all cases, expression ratios as determined by the two procedures differed by <2-fold for the genes identified in the heat shock experiments (Table 2). The two assays differed more widely in terms of assessing absolute expression levels; nonetheless, absolute expression as monitored on a microarray typically correlated with RNA blots to within a factor of five (Table 2).

Phorbol Ester Signaling. To explore a signaling pathway distinct from the heat shock response, microarrays were used to examine the cellular effects of phorbol ester treatment. Jurkat cells were treated with phorbol ester, harvested, lysed, and used as a source of mRNA. Samples of mRNA from untreated or phorbol ester-stimulated cells were labeled with reverse transcriptase. The probes were mixed, hybridized to microarrays, and scanned for fluorescence emission of the two fluorescent groups. A total of six array elements displayed ≥ 2.0 -fold elevated signals with probes from phorbol ester-treated cells relative to control samples (Fig. 2B).

To determine the identity of the phorbol ester-induced genes, clones corresponding to the six array elements were sequenced. Data base searches revealed perfect matches for five of the six sequences (Table 1). The two most highly induced genes were the PAC-1 tyrosine phosphatase and nuclear factor-kappa B1 (NF- κ B1); modest activation was observed for phosphoglycerate kinase and β_2 -microglobulin (Table 1). One remaining clone (B19) did not match any entry in the public data base (Table 1). B19 displayed a 2.1-fold induction and, similar to the novel heat shock genes, a relatively low absolute expression level (Tables 1 and 2). All six of the phorbol ester-inducible genes displayed increased steady-state mRNA levels by RNA blotting (Table 2). PAC-1 expression (Fig. 1; Table 2) defined a detection limit of ~ 0.2 copies per cell.

Transcript imaging of human tissue. To determine whether microarrays could be used to monitor expression in human tissues, probes were prepared from human bone mar-

row, brain, prostate, and heart by labeling each mRNA sample with Cy5-dCTP. In a separate reaction, a control probe was prepared by labeling Jurkat mRNA with fluorescein-dCTP. The four Cy5-labeled probes were each mixed with an aliquot of the fluorescein-labeled control sample, and the four mixtures were hybridized to separate microarrays. The arrays were washed and scanned for fluorescence emission, and hybridization signals for each of the tissues samples were normalized to the Jurkat control to generate an expression profile for each of the 1046 clones present on the array.

Detectable expression was observed for all 15 of the heat shock and phorbol ester-regulated genes in the four tissue types examined (Fig. 3). In general, the expression level of each gene in Jurkat cells correlated rather closely with expression in the four tissues (Table 2; Fig. 3). Genes encoding β -actin and cytochrome c oxidase, the two most highly expressed of the 15 genes in Jurkat cells (Table 2), were highly expressed in bone marrow, brain, prostate, and heart (Fig. 3A). Expression of cytochrome c oxidase, hsp90 α , and the novel B7 sequence was significantly greater in heart than in the other tissues (Fig. 3).

DISCUSSION

Many of the heat shock genes identified in this study encode factors that function either as molecular "chaperones" (HSP90 α , HSP90 β , DnaJ, TCP-1) or as mediators of protein degradation (polyubiquitin). The identification of these sequences is consistent with the biochemical basis of heat shock induction (10-15). Proteins undergo denaturation at elevated temperatures, and those that fail to maintain proper conformation must be selectively degraded (10-15). It will be interesting to determine whether the three novel heat shock genes identified in this study are involved in protein folding or degradation. The identification of these genes by expression monitoring, and biochemical analysis should provide a detailed functional understanding of these genes.

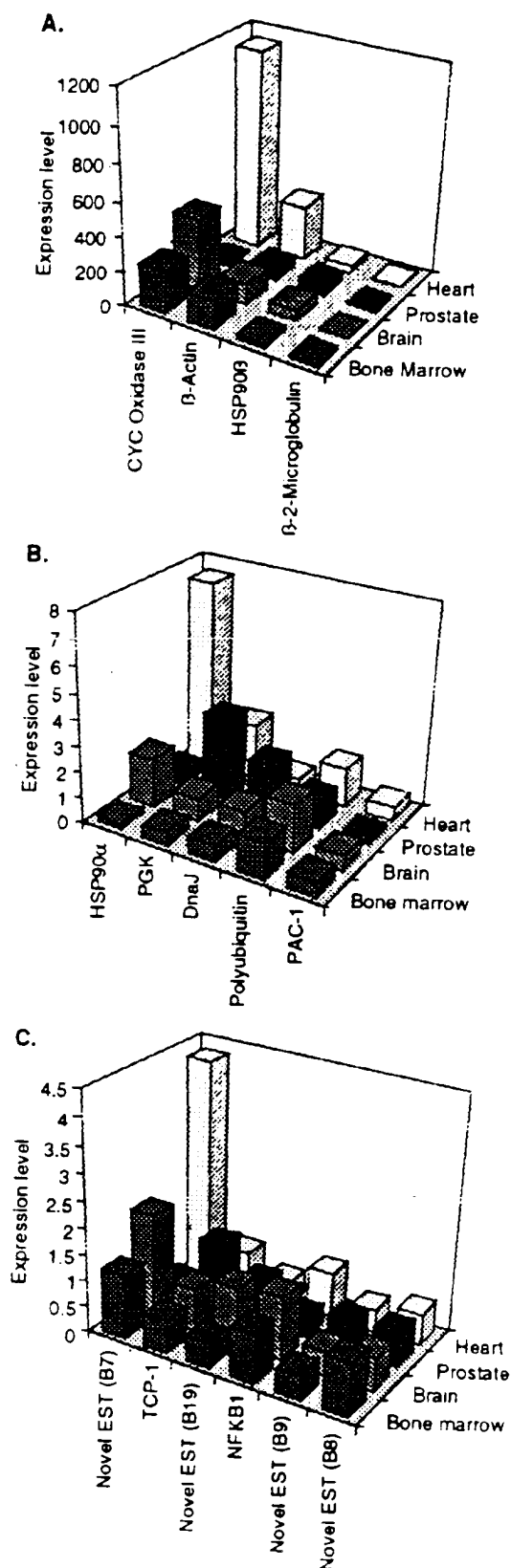


Fig. 4. Expression of genes in heart, prostate, brain, and bone marrow. The expression levels of genes in heart, prostate, brain, and bone marrow are shown. The genes are grouped according to expression levels (4–7).

Phorbol ester, a potent activator of protein kinase C (16, 17), induced a set of genes distinct from those involved in the heat shock pathway. The most highly induced gene identified in this study, *PAC-1*, encodes a nuclear tyrosine kinase that may play a role in regulating transcription and cell cycle progression (18). *NF-κB1*, a second phorbol ester-inducible gene, is an intensively studied member of the Rel transcription factor family (19–21). The Rel proteins are activated by a large number of stimuli, including phorbol esters, cytokines, bacterial and viral pathogens, and ultraviolet light (19–21). Modest activation was observed for three sequences not known to be inducible by phorbol esters, including phosphoglycerate kinase, β -microglobulin, and a novel human gene (B19). Extensive expression monitoring with microarrays should assist in understanding how each of these genes integrate into the highly complex phorbol ester signaling pathway.

It is striking that four novel human genes were discovered with an array of 1000 randomly chosen clones, particularly because the heat shock and phorbol ester signaling pathways have been so intensively studied (10–21). The facile discovery of these sequences underscores the fact that microarrays can be used for gene discovery in the absence of any sequence information. By this approach, clones are chosen at random from any library of interest and only those clones that display interesting expression patterns are sequenced and characterized. This parallel assay, coupled with a modest DNA sequencing facility, allows high-throughput human genome expression analysis and gene discovery.

Genes that are activated or repressed by a given stimulus provide functional clues to the cellular pathway involved (22–24). Detailed examination of these gene expression “signatures” can provide a dynamic view of the mode of action of a given signaling substance (22–24). Microarrays may thus allow rapid mechanistic examination of hormones, drugs, elicitors, and other small molecules; moreover, functional analysis of transcription factors, kinases, growth factors, cytokines, receptors, and other gene products should be possible. Efforts are underway to develop mRNA amplification strategies to enable probe preparation from minute tissue samples. This capability might allow for high-throughput patient screening in a clinical setting.

The current detection limit of the assay allows monitoring of transcripts that represent $\sim 1:500,000$ (wt/wt) of the total mRNA. This 10-fold increase in sensitivity compared with the original report (4) was achieved largely by modifying the coupling chemistry, which reduced background fluorescence. The significance of this improvement is considerable in that approximately half the human genes identified in this study, including all four novel sequences, exhibited expression levels below the original detection limit of 1:50,000 (4).

The ability to detect 2-fold changes in expression was achieved by the use of two-color fluorescence in the labeling and detection schemes, digitized data collection, and custom software. The importance of this capability is underscored by the fact that nearly all of the genes examined here exhibited <6 -fold changes in expression. The four novel genes, which showed ≤ 2.2 -fold activation, were probably overlooked in previous screens that used conventional differential expression techniques. It may be possible to further improve the precision of the microarray assay by the use of closely related fluorescent analogs, such as Cy3 and Cy5, in the labeling and hybridization reactions.

Microarrays offer a number of advantages over other potential high-capacity approaches to expression analysis. The chip-based approach enables small hybridization volumes, high

throughput, and the ability to use a wide variety of labeling and detection approaches (25, 26). The use of cDNA clones provides hybridization specificity that is not readily attained with oligo-

nucleotide arrays (27–30). The parallel format of the assay provides a simultaneous differential expression readout for >1000 genes. This contrasts with sequencing-based methods, which require serial data collection for expression analysis (31, 32). A commercial source of cDNA microarrays would greatly speed the use of a chip-based approach to expression analysis.

The availability of large numbers of ESTs (3) provides a rich resource of human cDNA clones for microarraying. The >400,000 ESTs in the public data bases represent a significant subset of all human genes (3, 33). Microarrays of thousands of ESTs will provide a powerful analytical tool for future human gene expression studies. The ~100,000 genes in the human genome (2, 33) emphasize the need for microarrays of greater density. Attempts to improve microdeposition techniques are underway and should allow construction of arrays containing a complete set of human gene targets (<http://cmgm.stanford.edu/~schena/>). Microarrays of ~100,000 cDNA elements would allow expression monitoring of the entire human genome in a single hybridization. This capacity, coupled with detailed biochemical analysis of the individual gene products, would greatly speed the functional analysis of the human genome.

We thank S. Elledge (selledge@bcm.tmc.edu) for the human cDNA library, Qiagen representatives for help with plasmid purification, and A. J. Smith and colleagues at the Protein and Nucleic Acid (PAN) facility (Stanford) for oligonucleotide synthesis and DNA sequencing. We also thank members of the Davis, Brown, and Smith laboratories for critical comments and helpful discussions and Synteni employees for technical assistance. Support for R.W.D. was provided by the National Science Foundation (MCB9106011) and National Institutes of Health (R37HG00198) and for P.O.B. by the National Institutes of Health (3R21HG00450) and Howard Hughes Medical Institute. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

- Watson, J. D. (1993) *Gene* 135, 309–315.
- Collins, F. S. (1995) *Proc. Natl. Acad. Sci. USA* 92, 10821–10823.
- Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R. & Venter, J. C. (1991) *Science* 252, 1651–1656.
- Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* 270, 467–470.
- Shalon, D. (1996) Ph.D. thesis (Stanford University).
- Schena, M. (1996) *BioEssays* 18, 427–431.
- Shalon, D., Smith, S. J. & Brown, P. O. (1996) *Genome Res.* 6, 639–645.
- Schena, M. & Davis, R. W. (1994) *Proc. Natl. Acad. Sci. USA* 91, 8393–8397.
- Schena, M. & Davis, R. W. (1992) *Proc. Natl. Acad. Sci. USA* 89, 3894–3898.
- Jindal, S. (1996) *Trends Biotechnol.* 14, 17–20.
- Wilkinson, K. D. (1995) *Annu. Rev. Nutr.* 15, 161–189.
- Jakob, U. & Buchner, J. (1994) *Trends Biochem. Sci.* 19, 205–211.
- Becker, J. & Craig, E. A. (1994) *Eur. J. Biochem.* 219, 11–23.
- Cyr, D. M., Langer, T. & Douglas, M. G. (1994) *Trends Biochem. Sci.* 19, 176–181.
- Craig, E. A., Weissman, J. S. & Horwich, A. L. (1994) *Cell* 78, 365–372.
- Newton, A. C. (1995) *J. Biol. Chem.* 270, 28495–28498.
- Nishizuka, Y. (1995) *FASEB J.* 9, 484–496.
- Rohan, P. J., Davis, P., Moskaluk, C. A., Kearns, M., Krutzsch, H., Siebenlist, U. & Kelly, K. (1993) *Science* 259, 1763–1766.
- Thanos, D. & Maniatis, T. (1995) *Cell* 80, 529–532.
- Baeuerle, P. A. & Henkel, T. (1994) *Annu. Rev. Immunol.* 12, 141–179.
- Liou, H.-C. & Baltimore, D. (1993) *Curr. Opin. Cell Biol.* 5, 477–487.
- Cohen, G. B., Ren, R. & Baltimore, D. (1995) *Cell* 80, 237–248.
- Chan, A. C., Desai, D. M. & Weiss, A. (1994) *Annu. Rev. Immunol.* 12, 555–592.
- Crabtree, G. R. & Clipstone, N. A. (1994) *Annu. Rev. Biochem.* 63, 1045–1083.
- Gress, T. M., Hoheisel, J. D., Lennon, G. G., Zehetner, G. & Lehrach, H. (1992) *Mamm. Genome* 3, 609–619.
- Bernard, K., Auphan, N., Granjeaud, S., Victorero, G., Schmitt-Verhulst, A.-M., Jordan, B. R. & Nguyen, C. (1996) *Nucleic Acids Res.* 24, 1435–1442.
- Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. (1991) *Science* 251, 767–773.
- Southern, E. M., Maskos, U. & Elder, J. K. (1992) *Genomics* 13, 1008–1017.
- Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. (1994) *Nucleic Acids Res.* 22, 5456–5465.
- Matson, R. S., Rampal, J., Pentoney, S. L., Jr., Anderson, P. D. & Coassin, P. (1995) *Anal. Biochem.* 224, 110–116.
- Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. (1995) *Science* 270, 484–487.
- Adams, M. D. (1996) *BioEssays* 18, 261–262.
- Fields, C., Adams, M. D., White, O. & Venter, J. C. (1994) *Nat. Genet.* 7, 345–346.

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G01N 33/543, 33/68	A1	(11) International Publication Number: WO 95/35505 (43) International Publication Date: 28 December 1995 (28.12.95)
(21) International Application Number: PCT/US95/07659 (22) International Filing Date: 16 June 1995 (16.06.95) (30) Priority Data: 08/261,388 17 June 1994 (17.06.94) US 08/477,809 7 June 1995 (07.06.95) US (71) Applicant: THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY [US/US]; Stanford, CA 94305 (US). (72) Inventors: SHALON, Tidhar, Dari; 364 Fletcher Drive, Atherton, CA 94027 (US). BROWN, Patrićk, O.; 76 Peter Coutts Circle, Stanford, CA 94305 (US). (74) Agent: DEHLINGER, Peter, J.; Dehlinger & Associates, P.O. Box 60850, Palo Alto, CA 94306-1546 (US).	(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>	
(54) Title: METHOD AND APPARATUS FOR FABRICATING MICROARRAYS OF BIOLOGICAL SAMPLES (57) Abstract A method and apparatus for forming microarrays of biological samples on a support are disclosed. The method involves dispensing a known volume of a reagent at each of a selected array position, by tapping a capillary dispenser on the support under conditions effective to draw a defined volume of liquid onto the support. The apparatus is designed to produce a microarray of such regions in an automated fashion.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain			US	United States of America
FI	Finland				

**METHOD AND APPARATUS FOR FABRICATING
MICROARRAYS OF BIOLOGICAL SAMPLES**

Field of the Invention

5 This invention relates to a method and apparatus
for fabricating microarrays of biological samples for
large scale screening assays, such as arrays of DNA
samples to be used in DNA hybridization assays for
genetic research and diagnostic applications.

10

References

- Abouzied, et al., *Journal of AOAC International*
77(2):495-500 (1994).
- Bohlander, et al., *Genomics* 13:1322-1324 (1992).
- 15 Drmanac, et al., *Science* 260:1649-1652 (1993).
- Fodor, et al., *Science* 251:767-773 (1991).
- Khrapko, et al., *DNA Sequence* 1:375-388 (1991).
- Kuriyama, et al., AN ISFET BIOSENSOR, APPLIED BIOSENSORS
(Donald Wise, Ed.), Butterworths, pp. 93-114 (1989).
- 20 Lehrach, et al., HYBRIDIZATION FINGERPRINTING IN GENOME
MAPPING AND SEQUENCING, GENOME ANALYSIS, VOL 1 (Davies and
Tilgham, Eds.), Cold Spring Harbor Press, pp. 39-81
(1990).
- Maniatis, et al., MOLECULAR CLONING, A LABORATORY
25 MANUAL, Cold Spring Harbor Press (1989).
- Nelson, et al., *Nature Genetics* 4:11-18 (1993).

Pirrung, et al., U.S. Patent No. 5,143,854 (1992).

Riles, et al., *Genetics* 134:81-150 (1993).

Schena, M. et al., *Proc. Nat. Acad. Sci. USA*
89:3894-3898 (1992).

5 Southern, et al., *Genomics* 13:1008-1017 (1992).

Background of the Invention

A variety of methods are currently available for making arrays of biological macromolecules, such as
10 arrays of nucleic acid molecules or proteins. One method for making ordered arrays of DNA on a porous membrane is a "dot blot" approach. In this method, a vacuum manifold transfers a plurality, e.g., 96, aqueous samples of DNA from 3 millimeter diameter wells
15 to a porous membrane. A common variant of this procedure is a "slot-blot" method in which the wells have highly-elongated oval shapes.

The DNA is immobilized on the porous membrane by baking the membrane or exposing it to UV radiation.
20 This is a manual procedure practical for making one array at a time and usually limited to 96 samples per array. "Dot-blot" procedures are therefore inadequate for applications in which many thousand samples must be determined.

25 A more efficient technique employed for making ordered arrays of genomic fragments uses an array of pins dipped into the wells, e.g., the 96 wells of a microtitre plate, for transferring an array of samples to a substrate, such as a porous membrane. One array
30 includes pins that are designed to spot a membrane in a staggered fashion, for creating an array of 9216 spots in a 22 x 22 cm area (Lehrach, et al., 1990). A limitation with this approach is that the volume of DNA spotted in each pixel of each array is highly variable.

In addition, the number of arrays that can be made with each dipping is usually quite small.

An alternate method of creating ordered arrays of nucleic acid sequences is described by Pirrung, et al. (1992), and also by Fodor, et al. (1991). The method involves synthesizing different nucleic acid sequences at different discrete regions of a support. This method employs elaborate synthetic schemes, and is generally limited to relatively short nucleic acid sample, e.g., less than 20 bases. A related method has been described by Southern, et al. (1992).

Khrapko, et al. (1991) describes a method of making an oligonucleotide matrix by spotting DNA onto a thin layer of polyacrylamide. The spotting is done manually with a micropipette.

None of the methods or devices described in the prior art are designed for mass fabrication of microarrays characterized by (i) a large number of micro-sized assay regions separated by a distance of 50-200 microns or less, and (ii) a well-defined amount, typically in the picomole range, of analyte associated with each region of the array.

Furthermore, current technology is directed at performing such assays one at a time to a single array of DNA molecules. For example, the most common method for performing DNA hybridizations to arrays spotted onto porous membrane involves sealing the membrane in a plastic bag (Maniatis, et al., 1989) or a rotating glass cylinder (Robbins Scientific) with the labeled hybridization probe inside the sealed chamber. For arrays made on non-porous surfaces, such as a microscope slide, each array is incubated with the labeled hybridization probe sealed under a coverslip. These techniques require a separate sealed chamber for

each array which makes the screening and handling of many such arrays inconvenient and time intensive.

Abouzied, et al. (1994) describes a method of printing horizontal lines of antibodies on a
5 nitrocellulose membrane and separating regions of the membrane with vertical stripes of a hydrophobic material. Each vertical stripe is then reacted with a different antigen and the reaction between the
10 immobilized antibody and an antigen is detected using a standard ELISA colorimetric technique. Abouzied's technique makes it possible to screen many one-dimensional arrays simultaneously on a single sheet of nitrocellulose. Abouzied makes the nitrocellulose
15 somewhat hydrophobic using a line drawn with PAP Pen (Research Products International). However Abouzied does not describe a technology that is capable of completely sealing the pores of the nitrocellulose. The pores of the nitrocellulose are still physically open and so the assay reagents can leak through the
20 hydrophobic barrier during extended high temperature incubations or in the presence of detergents which makes the Abouzied technique unacceptable for DNA hybridization assays.

Porous membranes with printed patterns of
25 hydrophilic/hydrophobic regions exist for applications such as ordered arrays of bacteria colonies. QA Life Sciences (San Diego CA) makes such a membrane with a grid pattern printed on it. However, this membrane has the same disadvantage as the Abouzied technique since
30 reagents can still flow between the gridded arrays making them unusable for separate DNA hybridization assays.

Pall Corporation make a 96-well plate with a porous filter heat sealed to the bottom of the plate.
35 These plates are capable of containing different

reagents in each well without cross-contamination. However, each well is intended to hold only one target element whereas the invention described here makes a microarray of many biomolecules in each subdivided region of the solid support. Furthermore, the 96 well plates are at least 1 cm thick and prevent the use of the device for many colorimetric, fluorescent and radioactive detection formats which require that the membrane lie flat against the detection surface. The invention described here requires no further processing after the assay step since the barriers elements are shallow and do not interfere with the detection step thereby greatly increasing convenience.

Hyseq Corporation has described a method of making an "array of arrays" on a non-porous solid support for use with their sequencing by hybridization technique. The method described by Hyseq involves modifying the chemistry of the solid support material to form a hydrophobic grid pattern where each subdivided region contains a microarray of biomolecules. Hyseq's flat hydrophobic pattern does not make use of physical blocking as an additional means of preventing cross contamination.

25 Summary of the Invention

The invention includes, in one aspect, a method of forming a microarray of analyte-assay regions on a solid support, where each region in the array has a known amount of a selected, analyte-specific reagent. The method involves first loading a solution of a selected analyte-specific reagent in a reagent-dispensing device having an elongate capillary channel (i) formed by spaced-apart, coextensive elongate members, (ii) adapted to hold a quantity of the reagent solution and (iii) having a tip region at which aqueous

solution in the channel forms a meniscus. The channel is preferably formed by a pair of spaced-apart tapered elements.

5 The tip of the dispensing device is tapped against a solid support at a defined position on the support surface with an impulse effective to break the meniscus in the capillary channel deposit a selected volume of solution on the surface, preferably a selected volume in the range 0.01 to 100 nl. The two steps are
10 repeated until the desired array is formed.

The method may be practiced in forming a plurality of such arrays, where the solution-depositing step is are applied to a selected position on each of a plurality of solid supports at each repeat cycle.

15 The dispensing device may be loaded with a new solution, by the steps of (i) dipping the capillary channel of the device in a wash solution, (ii) removing wash solution drawn into the capillary channel, and (iii) dipping the capillary channel into the new
20 reagent solution.

Also included in the invention is an automated apparatus for forming a microarray of analyte-assay regions on a plurality of solid supports, where each region in the array has a known amount of a selected,
25 analyte-specific reagent. The apparatus has a holder for holding, at known positions, a plurality of planar supports, and a reagent dispensing device of the type described above.

The apparatus further includes positioning
30 structure for positioning the dispensing device at a selected array position with respect to a support in said holder, and dispensing structure for moving the dispensing device into tapping engagement against a support with a selected impulse effective to deposit a

selected volume on the support, e.g., a selected volume in the volume range 0.01 to 100 nl.

The positioning and dispensing structures are controlled by a control unit in the apparatus. The unit operates to (i) place the dispensing device at a loading station, (ii) move the capillary channel in the device into a selected reagent at the loading station, to load the dispensing device with the reagent, and (iii) dispense the reagent at a defined array position on each of the supports on said holder. The unit may further operate, at the end of a dispensing cycle, to wash the dispensing device by (i) placing the dispensing device at a washing station, (ii) moving the capillary channel in the device into a wash fluid, to load the dispensing device with the fluid, and (iii) remove the wash fluid prior to loading the dispensing device with a fresh selected reagent.

The dispensing device in the apparatus may be one of a plurality of such devices which are carried on the arm for dispensing different analyte assay reagents at selected spaced array positions.

In another aspect, the invention includes a substrate with a surface having a microarray of at least 10^3 distinct polynucleotide or polypeptide biopolymers in a surface area of less than about 1 cm^2 . Each distinct biopolymer (i) is disposed at a separate, defined position in said array, (ii) has a length of at least 50 subunits, and (iii) is present in a defined amount between about 0.1 femtomoles and 100 nanomoles.

In one embodiment, the surface is glass slide surface coated with a polycationic polymer, such as polylysine, and the biopolymers are polynucleotides. In another embodiment, the substrate has a water-impermeable backing, a water-permeable film formed on

the backing, and a grid formed on the film. The grid is composed of intersecting water-impervious grid elements extending from said backing to positions raised above the surface of said film, and partitions the film into a plurality of water-impervious cells. A biopolymer array is formed within each well.

More generally, there is provided a substrate for use in detecting binding of labeled polynucleotides to one or more of a plurality different-sequence, immobilized polynucleotides. The substrate includes, in one aspect, a glass support, a coating of a polycationic polymer, such as polylysine, on said surface of the support, and an array of distinct polynucleotides electrostatically bound non-covalently to said coating, where each distinct biopolymer is disposed at a separate, defined position in a surface array of polynucleotides.

In another aspect, the substrate includes a water-impermeable backing, a water-permeable film formed on the backing, and a grid formed on the film, where the grid is composed of intersecting water-impervious grid elements extending from the backing to positions raised above the surface of the film, forming a plurality of cells. A biopolymer array is formed within each cell.

Also forming part of the invention is a method of detecting differential expression of each of a plurality of genes in a first cell type, with respect to expression of the same genes in a second cell type. In practicing the method, there is first produced fluorescent-labeled cDNA's from mRNA's isolated from the two cells types, where the cDNA'S from the first and second cells are labeled with first and second different fluorescent reporters.

A mixture of the labeled cDNA's from the two cell types is added to an array of polynucleotides

representing a plurality of known genes derived from the two cell types, under conditions that result in hybridization of the cDNA's to complementary-sequence polynucleotides in the array. The array is then
5 examined by fluorescence under fluorescence excitation conditions in which (i) polynucleotides in the array that are hybridized predominantly to cDNA's derived from one of the first and second cell types give a distinct first or second fluorescence emission color,
10 respectively, and (ii) polynucleotides in the array that are hybridized to substantially equal numbers of cDNA's derived from the first and second cell types give a distinct combined fluorescence emission color, respectively. The relative expression of known genes
15 in the two cell types can then be determined by the observed fluorescence emission color of each spot.

These and other objects and features of the invention will become more fully apparent when the following detailed description of the invention is read
20 in conjunction with the accompanying figures.

Brief Description of the Drawings

Fig. 1 is a side view of a reagent-dispensing device having an open-capillary dispensing head
25 constructed for use in one embodiment of the invention;

Figs. 2A-2C illustrate steps in the delivery of a fixed-volume bead on a hydrophobic surface employing the dispensing head from Fig. 1, in accordance with one embodiment of the method of the invention;

30 Fig. 3 shows a portion of a two-dimensional array of analyte-assay regions constructed according to the method of the invention;

Fig. 4 is a planar view showing components of an automated apparatus for forming arrays in accordance
35 with the invention.

Fig. 5 shows a fluorescent image of an actual 20 × 20 array of 400 fluorescently-labeled DNA samples immobilized on a poly-l-lysine coated slide, where the total area covered by the 400 element array is 16 square millimeters;

Fig. 6 is a fluorescent image of a 1.8 cm × 1.8 cm microarray containing lambda clones with yeast inserts, the fluorescent signal arising from the hybridization to the array with approximately half the yeast genome labeled with a green fluorophore and the other half with a red fluorophore;

Fig. 7 shows the translation of the hybridization image of Fig. 6 into a karyotype of the yeast genome, where the elements of Fig.-6 microarray contain yeast DNA sequences that have been previously physically mapped in the yeast genome;

Fig. 8 show a fluorescent image of a 0.5 cm × 0.5 cm microarray of 24 cDNA clones, where the microarray was hybridized simultaneously with total cDNA from wild type *Arabidopsis* plant labeled with a green fluorophore and total cDNA from a transgenic *Arabidopsis* plant labeled with a red fluorophore, and the arrow points to the cDNA clone representing the gene introduced into the transgenic *Arabidopsis* plant;

Fig. 9 shows a plan view of substrate having an array of cells formed by barrier elements in the form of a grid;

Fig. 10 shows an enlarged plan view of one of the cells in the substrate in Fig. 9, showing an array of polynucleotide regions in the cell;

Fig. 11 is an enlarged sectional view of the substrate in Fig. 9, taken along a section line in that figure; and

Fig. 12 is a scanned image of a 3 cm × 3 cm nitrocellulose solid support containing four identical

arrays of M13 clones in each of four quadrants, where each quadrant was hybridized simultaneously to a different oligonucleotide using an open face hybridization method.

5

Detailed Description of the Invention

I. Definitions

Unless indicated otherwise, the terms defined below have the following meanings:

10 "Ligand" refers to one member of a ligand/anti-ligand binding pair. The ligand may be, for example, one of the nucleic acid strands in a complementary, hybridized nucleic acid duplex binding pair; an effector molecule in an effector/receptor binding pair;
15 or an antigen in an antigen/antibody or antigen/antibody fragment binding pair.

"Antiligand" refers to the opposite member of a ligand/anti-ligand binding pair. The antiligand may be the other of the nucleic acid strands in a
20 complementary, hybridized nucleic acid duplex binding pair; the receptor molecule in an effector/receptor binding pair; or an antibody or antibody fragment molecule in antigen/antibody or antigen/antibody fragment binding pair, respectively.

25 "Analyte" or "analyte molecule" refers to a molecule, typically a macromolecule, such as a polynucleotide or polypeptide, whose presence, amount, and/or identity are to be determined. The analyte is one member of a ligand/anti-ligand pair.

30 "Analyte-specific assay reagent" refers to a molecule effective to bind specifically to an analyte molecule. The reagent is the opposite member of a ligand/anti-ligand binding pair.

An "array of regions on a solid support" is a
35 linear or two-dimensional array of preferably discrete

regions, each having a finite area, formed on the surface of a solid support.

A "microarray" is an array of regions having a density of discrete regions of at least about $100/\text{cm}^2$, and preferably at least about $1000/\text{cm}^2$. The regions in a microarray have typical dimensions, e.g., diameters, in the range of between about $10\text{-}250\text{ }\mu\text{m}$, and are separated from other regions in the array by about the same distance.

A support surface is "hydrophobic" if a aqueous-medium droplet applied to the surface does not spread out substantially beyond the area size of the applied droplet. That is, the surface acts to prevent spreading of the droplet applied to the surface by hydrophobic interaction with the droplet.

A "meniscus" means a concave or convex surface that forms on the bottom of a liquid in a channel as a result of the surface tension of the liquid.

"Distinct biopolymers", as applied to the biopolymers forming a microarray, means an array member which is distinct from other array members on the basis of a different biopolymer sequence, and/or different concentrations of the same or distinct biopolymers, and/or different mixtures of distinct or different-concentration biopolymers. Thus an array of "distinct polynucleotides" means an array containing, as its members, (i) distinct polynucleotides, which may have a defined amount in each member, (ii) different, graded concentrations of given-sequence polynucleotides, and/or (iii) different-composition mixtures of two or more distinct polynucleotides.

"Cell type" means a cell from a given source, e.g., a tissue, or organ, or a cell in a given state of

differentiation, or a cell associated with a given pathology or genetic makeup.

II. Method of Microarray Formation

5 This section describes a method of forming a microarray of analyte-assay regions on a solid support or substrate, where each region in the array has a known amount of a selected, analyte-specific reagent.

10 Fig. 1 illustrates, in a partially schematic view, a reagent-dispensing device 10 useful in practicing the method. The device generally includes a reagent dispenser 12 having an elongate open capillary channel 14 adapted to hold a quantity of the reagent solution, such as indicated at 16, as will be described below.

15 The capillary channel is formed by a pair of spaced-apart, coextensive, elongate members 12a, 12b which are tapered toward one another and converge at a tip or tip region 18 at the lower end of the channel. More generally, the open channel is formed by at least two

20 elongate, spaced-apart members adapted to hold a quantity of reagent solutions and having a tip region at which aqueous solution in the channel forms a meniscus, such as the concave meniscus illustrated at 20 in Fig. 2A. The advantages of the open channel

25 construction of the dispenser are discussed below.

 With continued reference to Fig. 1, the dispenser device also includes structure for moving the dispenser rapidly toward and away from a support surface, for effecting deposition of a known amount of solution in

30 the dispenser on a support, as will be described below with reference to Figs. 2A-2C. In the embodiment shown, this structure includes a solenoid 22 which is activatable to draw a solenoid piston 24 rapidly downwardly, then release the piston, e.g., under spring

35 bias, to a normal, raised position, as shown. The

dispenser is carried on the piston by a connecting member 26, as shown. The just-described moving structure is also referred to herein as dispensing means for moving the dispenser into engagement with a solid support, for dispensing a known volume of fluid on the support.

The dispensing device just described is carried on an arm 28 that may be moved either linearly or in an x-y plane to position the dispenser at a selected deposition position, as will be described.

Figs. 2A-2C illustrate the method of depositing a known amount of reagent solution in the just-described dispenser on the surface of a solid support, such as the support indicated at 30. The support is a polymer, glass, or other solid-material support having a surface indicated at 31.

In one general embodiment, the surface is a relatively hydrophilic, i.e., wettable surface, such as a surface having native, bound or covalently attached charged groups. On such surface described below is a glass surface having an absorbed layer of a polycationic polymer, such as poly-l-lysine.

In another embodiment, the surface has or is formed to have a relatively hydrophobic character, i.e., one that causes aqueous medium deposited on the surface to bead. A variety of known hydrophobic polymers, such as polystyrene, polypropylene, or polyethylene have desired hydrophobic properties, as do glass and a variety of lubricant or other hydrophobic films that may be applied to the support surface.

Initially, the dispenser is loaded with a selected analyte-specific reagent solution, such as by dipping the dispenser tip, after washing, into a solution of the reagent, and allowing filling by capillary flow into the dispenser channel. The dispenser is now moved

to a selected position with respect to a support surface, placing the dispenser tip directly above the support-surface position at which the reagent is to be deposited. This movement takes place with the
5 dispenser tip in its raised position, as seen in Fig. 2A, where the tip is typically at least several 1-5 mm above the surface of the substrate.

With the dispenser so positioned, solenoid 22 is now activated to cause the dispenser tip to move
10 rapidly toward and away from the substrate surface, making momentary contact with the surface, in effect, tapping the tip of the dispenser against the support surface. The tapping movement of the tip against the surface acts to break the liquid meniscus in the tip
15 channel, bringing the liquid in the tip into contact with the support surface. This, in turn, produces a flowing of the liquid into the capillary space between the tip and the surface, acting to draw liquid out of the dispenser channel, as seen in Fig. 2B.

20 Fig. 2C shows flow of fluid from the tip onto the support surface, which in this case is a hydrophobic surface. The figure illustrates that liquid continues to flow from the dispenser onto the support surface until it forms a liquid bead 32. At a given bead size,
25 i.e., volume, the tendency of liquid to flow onto the surface will be balanced by the hydrophobic surface interaction of the bead with the support surface, which acts to limit the total bead area on the surface, and by the surface tension of the droplet, which tends
30 toward a given bead curvature. At this point, a given bead volume will have formed, and continued contact of the dispenser tip with the bead, as the dispenser tip is being withdrawn, will have little or no effect on bead volume.

For liquid-dispensing on a more hydrophilic surface, the liquid will have less of a tendency to bead, and the dispensed volume will be more sensitive to the total dwell time of the dispenser tip in the immediate vicinity of the support surface, e.g., the positions illustrated in Figs. 2B and 2C.

The desired deposition volume, i.e., bead volume, formed by this method is preferably in the range 2 pl (picoliters) to 2 nl (nanoliters), although volumes as high as 100 nl or more may be dispensed. It will be appreciated that the selected dispensed volume will depend on (i) the "footprint" of the dispenser tip, i.e., the size of the area spanned by the tip, (ii) the hydrophobicity of the support surface, and (iii) the time of contact with and rate of withdrawal of the tip from the support surface. In addition, bead size may be reduced by increasing the viscosity of the medium, effectively reducing the flow time of liquid from the dispenser onto the support surface. The drop size may be further constrained by depositing the drop in a hydrophilic region surrounded by a hydrophobic grid pattern on the support surface.

In a typical embodiment, the dispenser tip is tapped rapidly against the support surface, with a total residence time in contact with the support of less than about 1 msec, and a rate of upward travel from the surface of about 10 cm/sec.

Assuming that the bead that forms on contact with the surface is a hemispherical bead, with a diameter approximately equal to the width of the dispenser tip, as shown in Fig. 2C, the volume of the bead formed in relation to dispenser tip width (d) is given in Table 1 below. As seen, the volume of the bead ranges between 2 pl to 2 nl as the width size is increased from about 20 to 200 μm .

Table 1

d	Volume (nl)
20 μm	2×10^{-3}
50 μm	3.1×10^{-2}
100 μm	2.5×10^{-1}
200 μm	2

5
10 At a given tip size, bead volume can be reduced in a controlled fashion by increasing surface hydrophobicity, reducing time of contact of the tip with the surface, increasing rate of movement of the tip away from the surface, and/or increasing the
15 viscosity of the medium. Once these parameters are fixed, a selected deposition volume in the desired pl to nl range can be achieved in a repeatable fashion.

After depositing a bead at one selected location on a support, the tip is typically moved to a
20 corresponding position on a second support, a droplet is deposited at that position, and this process is repeated until a liquid droplet of the reagent has been deposited at a selected position on each of a plurality of supports.

25 The tip is then washed to remove the reagent liquid, filled with another reagent liquid and this reagent is now deposited at each another array position on each of the supports. In one embodiment, the tip is washed and refilled by the steps of (i) dipping the
30 capillary channel of the device in a wash solution, (ii) removing wash solution drawn into the capillary channel, and (iii) dipping the capillary channel into the new reagent solution.

From the foregoing, it will be appreciated that
35 the tweezers-like, open-capillary dispenser tip

provides the advantages that (i) the open channel of the tip facilitates rapid, efficient washing and drying before reloading the tip with a new reagent, (ii) passive capillary action can load the sample directly from a standard microwell plate while retaining sufficient sample in the open capillary reservoir for the printing of numerous arrays, (iii) open capillaries are less prone to clogging than closed capillaries, and (iv) open capillaries do not require a perfectly faced bottom surface for fluid delivery.

A portion of a microarray 36 formed on the surface 38 of a solid support 40 in accordance with the method just described is shown in Fig. 3. The array is formed of a plurality of analyte-specific reagent regions, such as regions 42, where each region may include a different analyte-specific reagent. As indicated above, the diameter of each region is preferably between about 20-200 μm . The spacing between each region and its closest (non-diagonal) neighbor, measured from center-to-center (indicated at 44), is preferably in the range of about 20-400 μm . Thus, for example, an array having a center-to-center spacing of about 250 μm contains about 40 regions/cm or 1,600 regions/cm². After formation of the array, the support is treated to evaporate the liquid of the droplet forming each region, to leave a desired array of dried, relatively flat regions. This drying may be done by heating or under vacuum.

In some cases, it is desired to first rehydrate the droplets containing the analyte reagents to allow for more time for adsorption to the solid support. It is also possible to spot out the analyte reagents in a humid environment so that droplets do not dry until the arraying operation is complete.

III. Automated Apparatus for Forming Arrays

In another aspect, the invention includes an automated apparatus for forming an array of analyte-assay regions on a solid support, where each region in the array has a known amount of a selected, analyte-specific reagent.

The apparatus is shown in planar, and partially schematic view in Fig. 4. A dispenser device 72 in the apparatus has the basic construction described above with respect to Fig. 1, and includes a dispenser 74 having an open-capillary channel terminating at a tip, substantially as shown in Figs. 1 and 2A-2C.

The dispenser is mounted in the device for movement toward and away from a dispensing position at which the tip of the dispenser taps a support surface, to dispense a selected volume of reagent solution, as described above. This movement is effected by a solenoid 76 as described above. Solenoid 76 is under the control of a control unit 77 whose operation will be described below. The solenoid is also referred to herein as dispensing means for moving the device into tapping engagement with a support, when the device is positioned at a defined array position with respect to that support.

The dispenser device is carried on an arm 74 which is threadedly mounted on a worm screw 80 driven (rotated) in a desired direction by a stepper motor 82 also under the control of unit 77. At its left end in the figure screw 80 is carried in a sleeve 84 for rotation about the screw axis. At its other end, the screw is mounted to the drive shaft of the stepper motor, which in turn is carried on a sleeve 86. The dispenser device, worm screw, the two sleeves mounting the worm screw, and the stepper motor used in moving the device in the "x" (horizontal) direction in the

figure form what is referred to here collectively as a displacement assembly 86.

The displacement assembly is constructed to produce precise, micro-range movement in the direction of the screw, *i.e.*, along an x axis in the figure. In one mode, the assembly functions to move the dispenser in x-axis increments having a selected distance in the range 5-25 μm . In another mode, the dispenser unit may be moved in precise x-axis increments of several microns or more, for positioning the dispenser at associated positions on adjacent supports, as will be described below.

The displacement assembly, in turn, is mounted for movement in the "y" (vertical) axis of the figure, for positioning the dispenser at a selected y axis position. The structure mounting the assembly includes a fixed rod 88 mounted rigidly between a pair of frame bars 90, 92, and a worm screw 94 mounted for rotation between a pair of frame bars 96, 98. The worm screw is driven (rotated) by a stepper motor 100 which operates under the control of unit 77. The motor is mounted on bar 96, as shown.

The structure just described, including worm screw 94 and motor 100, is constructed to produce precise, micro-range movement in the direction of the screw, *i.e.*, along an y axis in the figure. As above, the structure functions in one mode to move the dispenser in y-axis increments having a selected distance in the range 5-250 μm , and in a second mode, to move the dispenser in precise y-axis increments of several microns (μm) or more, for positioning the dispenser at associated positions on adjacent supports.

The displacement assembly and structure for moving this assembly in the y axis are referred to herein collectively as positioning means for positioning the

dispensing device at a selected array position with respect to a support.

A holder 102 in the apparatus functions to hold a plurality of supports, such as supports 104 on which the microarrays of reagent regions are to be formed by the apparatus. The holder provides a number of recessed slots, such as slot 106, which receive the supports, and position them at precise selected positions with respect to the frame bars on which the dispenser moving means is mounted.

As noted above, the control unit in the device functions to actuate the two stepper motors and dispenser solenoid in a sequence designed for automated operation of the apparatus in forming a selected microarray of reagent regions on each of a plurality of supports.

The control unit is constructed, according to conventional microprocessor control principles, to provide appropriate signals to each of the solenoid and each of the stepper motors, in a given timed sequence and for appropriate signalling time. The construction of the unit, and the settings that are selected by the user to achieve a desired array pattern, will be understood from the following description of a typical apparatus operation.

Initially, one or more supports are placed in one or more slots in the holder. The dispenser is then moved to a position directly above a well (not shown) containing a solution of the first reagent to be dispensed on the support(s). The dispenser solenoid is actuated now to lower the dispenser tip into this well, causing the capillary channel in the dispenser to fill. Motors 82, 100 are now actuated to position the dispenser at a selected array position at the first of the supports. Solenoid actuation of the dispenser is

then effective to dispense a selected-volume droplet of that reagent at this location. As noted above, this operation is effective to dispense a selected volume preferably between 2 pl and 2 nl of the reagent
5 solution.

The dispenser is now moved to the corresponding position at an adjacent support and a similar volume of the solution is dispensed at this position. The process is repeated until the reagent has been
10 dispensed at this preselected corresponding position on each of the supports.

Where it is desired to dispense a single reagent at more than two array positions on a support, the dispenser may be moved to different array positions at
15 each support, before moving the dispenser to a new support, or solution can be dispensed at individual positions on each support, at one selected position, then the cycle repeated for each new array position.

To dispense the next reagent, the dispenser is
20 positioned over a wash solution (not shown), and the dispenser tip is dipped in and out of this solution until the reagent solution has been substantially washed from the tip. Solution can be removed from the tip, after each dipping, by vacuum, compressed air
25 spray, sponge, or the like.

The dispenser tip is now dipped in a second reagent well, and the filled tip is moved to a second selected array position in the first support. The process of dispensing reagent at each of the
30 corresponding second-array positions is then carried as above. This process is repeated until an entire microarray of reagent solutions on each of the supports has been formed.

35 IV. Microarray Substrate

This section describes embodiments of a substrate having a microarray of biological polymers carried on the substrate surface. Subsection A describes a multi-cell substrate, each cell of which contains a microarray, and preferably an identical microarray, of distinct biopolymers, such as distinct polynucleotides, formed on a porous surface. Subsection B describes a microarray of distinct polynucleotides bound on a glass slide coated with a polycationic polymer.

A. Multi-Cell Substrate

Fig. 9 illustrates, in plan view, a substrate 110 constructed according to the invention. The substrate has an 8 x 12 rectangular array 112 of cells, such as cells 114, 116, formed on the substrate surface. With reference to Fig. 10, each cell, such as cell 114, in turn supports a microarray 118 of distinct biopolymers, such as polypeptides or polynucleotides at known, addressable regions of the microarray. Two such regions forming the microarray are indicated at 120, and correspond to regions, such as regions 42, forming the microarray of distinct biopolymers shown in Fig. 3.

The 96-cell array shown in Fig. 9 has typically array dimensions between about 12 and 244 mm in width and 8 and 400 mm in length, with the cells in the array having width and length dimension of 1/12 and 1/8 the array width and length dimensions, respectively, i.e., between about 1 and 20 in width and 1 and 50 mm in length.

The construction of substrate is shown cross-sectionally in Fig. 11, which is an enlarged sectional view taken along view line 124 in Fig. 9. The substrate includes a water-impermeable backing 126, such as a glass slide or rigid polymer sheet. Formed on the surface of the backing is a water-permeable film

128. The film is formed of a porous membrane material, such as nitrocellulose membrane, or a porous web material, such as a nylon, polypropylene, or PVDF porous polymer material. The thickness of the film is preferably between about 10 and 1000 μm . The film may be applied to the backing by spraying or coating uncured material on the backing, or by applying a preformed membrane to the backing. The backing and film may be obtained as a preformed unit from commercial source, e.g., a plastic-backed nitrocellulose film available from Schleicher and Schuell Corporation.

With continued reference to Fig. 11, the film-covered surface in the substrate is partitioned into a desired array of cells by water-impermeable grid lines, such as lines 130, 132, which have infiltrated the film down to the level of the backing, and extend above the surface of the film as shown, typically a distance of 100 to 2000 μm above the film surface.

The grid lines are formed on the substrate by laying down an uncured or otherwise flowable resin or elastomer solution in an array grid, allowing the material to infiltrate the porous film down to the backing, then curing or otherwise hardening the grid lines to form the cell-array substrate.

One preferred material for the grid is a flowable silicone available from Loctite Corporation. The barrier material can be extruded through a narrow syringe (e.g., 22 gauge) using air pressure or mechanical pressure. The syringe is moved relative to the solid support to print the barrier elements as a grid pattern. The extruded bead of silicone wicks into the pores of the solid support and cures to form a shallow waterproof barrier separating the regions of the solid support.

In alternative embodiments, the barrier element can be a wax-based material or a thermoset material such as epoxy. The barrier material can also be a UV-curing polymer which is exposed to UV light after being printed onto the solid support. The barrier material may also be applied to the solid support using printing techniques such as silk-screen printing. The barrier material may also be a heat-seal stamping of the porous solid support which seals its pores and forms a water-impervious barrier element. The barrier material may also be a shallow grid which is laminated or otherwise adhered to the solid support.

In addition to plastic-backed nitrocellulose, the solid support can be virtually any porous membrane with or without a non-porous backing. Such membranes are readily available from numerous vendors and are made from nylon, PVDF, polysulfone and the like. In an alternative embodiment, the barrier element may also be used to adhere the porous membrane to a non-porous backing in addition to functioning as a barrier to prevent cross contamination of the assay reagents.

In an alternative embodiment, the solid support can be of a non-porous material. The barrier can be printed either before or after the microarray of biomolecules is printed on the solid support.

As can be appreciated, the cells formed by the grid lines and the underlying backing are water-impermeable, having side barriers projecting above the porous film in the cells. Thus, defined-volume samples can be placed in each well without risk of cross-contamination with sample material in adjacent cells. In Fig. 11, defined volumes samples, such as sample 134, are shown in the cells.

As noted above, each well contains a microarray of distinct biopolymers. In one general embodiment, the

microarrays in the well are identical arrays of distinct biopolymers, e.g., different sequence polynucleotides. Such arrays can be formed in accordance with the methods described in Section II, by
5 depositing a first selected polynucleotide at the same selected microarray position in each of the cells, then depositing a second polynucleotide at a different microarray position in each well, and so on until a complete, identical microarray is formed in each cell.

10 In a preferred embodiment, each microarray contains about 10^3 distinct polynucleotide or polypeptide biopolymers per surface area of less than about 1 cm^2 . Also in a preferred embodiment, the biopolymers in each microarray region are present in a
15 defined amount between about 0.1 femtomoles and 100 nanomoles. The ability to form high-density arrays of biopolymers, where each region is formed of a well-defined amount of deposited material, can be achieved in accordance with the microarray-forming method
20 described in Section II.

Also in a preferred embodiments, the biopolymers are polynucleotides having lengths of at least about 50 bp, i.e., substantially longer than oligonucleotides which can be formed in high-density arrays by schemes
25 involving parallel, step-wise polymer synthesis on the array surface.

In the case of a polynucleotide array, in an assay procedure, a small volume of the labeled DNA probe mixture in a standard hybridization solution is loaded
30 onto each cell. The solution will spread to cover the entire microarray and stop at the barrier elements. The solid support is then incubated in a humid chamber at the appropriate temperature as required by the assay.

Each assay may be conducted in an "open-face" format where no further sealing step is required, since the hybridization solution will be kept properly hydrated by the water vapor in the humid chamber. At the conclusion of the incubation step, the entire solid support containing the numerous microarrays is rinsed quickly enough to dilute the assay reagents so that no significant cross contamination occurs. The entire solid support is then reacted with detection reagents if needed and analyzed using standard colorimetric, radioactive or fluorescent detection means. All processing and detection steps are performed simultaneously to all of the microarrays on the solid support ensuring uniform assay conditions for all of the microarrays on the solid support.

B. Glass-Slide Polynucleotide Array

Fig. 5 shows a substrate 136 formed according to another aspect of the invention, and intended for use in detecting binding of labeled polynucleotides to one or more of a plurality distinct polynucleotides. The substrate includes a glass substrate 138 having formed on its surface, a coating of a polycationic polymer, preferably a cationic polypeptide, such as polylysine or polyarginine. Formed on the polycationic coating is a microarray 140 of distinct polynucleotides, each localized at known selected array regions, such as regions 142.

The slide is coated by placing a uniform-thickness film of a polycationic polymer, e.g., poly-l-lysine, on the surface of a slide and drying the film to form a dried coating. The amount of polycationic polymer added is sufficient to form at least a monolayer of polymers on the glass surface. The polymer film is bound to surface via electrostatic binding between

negative silyl-OH groups on the surface and charged amine groups in the polymers. Poly-L-lysine coated glass slides may be obtained commercially, e.g., from Sigma Chemical Co. (St. Louis, MO).

5 To form the microarray, defined volumes of distinct polynucleotides are deposited on the polymer-coated slide, as described in Section II. According to an important feature of the substrate, the deposited polynucleotides remain bound to the coated slide
10 surface non-covalently when an aqueous DNA sample is applied to the substrate under conditions which allow hybridization of reporter-labeled polynucleotides in the sample to complementary-sequence (single-stranded) polynucleotides in the substrate array. The method is
15 illustrated in Examples 1 and 2.

To illustrate this feature, a substrate of the type just described, but having an array of same-sequence polynucleotides, was mixed with fluorescent-labeled complementary DNA under hybridization
20 conditions. After washing to remove non-hybridized material, the substrate was examined by low-power fluorescence microscopy. The array can be visualized by the relatively uniform labeling pattern of the array regions.

25 In a preferred embodiment, each microarray contains at least 10^3 distinct polynucleotide or polypeptide biopolymers per surface area of less than about 1 cm^2 . In the embodiment shown in Fig. 5, the microarray contains 400 regions in an area of about 16 mm^2 , or 2.5×10^3 regions/ cm^2 . Also in a preferred
30 embodiment, the polynucleotides in the each microarray region are present in a defined amount between about 0.1 femtomoles and 100 nanomoles in the case of polynucleotides. As above, the ability to form high-

density arrays of this type, where each region is formed of a well-defined amount of deposited material, can be achieved in accordance with the microarray-forming method described in Section II.

5 Also in a preferred embodiments, the polynucleotides have lengths of at least about 50 bp, i.e., substantially longer than oligonucleotides which can be formed in high-density arrays by various in situ synthesis schemes.

10

V. Utility

Microarrays of immobilized nucleic acid sequences prepared in accordance with the invention can be used for large scale hybridization assays in numerous
15 genetic applications, including genetic and physical mapping of genomes, monitoring of gene expression, DNA sequencing, genetic diagnosis, genotyping of organisms, and distribution of DNA reagents to researchers.

For gene mapping, a gene or a cloned DNA fragment
20 is hybridized to an ordered array of DNA fragments, and the identity of the DNA elements applied to the array is unambiguously established by the pixel or pattern of pixels of the array that are detected. One application of such arrays for creating a genetic map is described
25 by Nelson, et al. (1993). In constructing physical maps of the genome, arrays of immobilized cloned DNA fragments are hybridized with other cloned DNA fragments to establish whether the cloned fragments in the probe mixture overlap and are therefore contiguous
30 to the immobilized clones on the array. For example, Lehrach, et al., describe such a process.

The arrays of immobilized DNA fragments may also be used for genetic diagnostics. To illustrate, an array containing multiple forms of a mutated gene or
35 genes can be probed with a labeled mixture of a

patient's DNA which will preferentially interact with only one of the immobilized versions of the gene.

The detection of this interaction can lead to a medical diagnosis. Arrays of immobilized DNA fragments can also be used in DNA probe diagnostics. For example, the identity of a pathogenic microorganism can be established unambiguously by hybridizing a sample of the unknown pathogen's DNA to an array containing many types of known pathogenic DNA. A similar technique can also be used for unambiguous genotyping of any organism. Other molecules of genetic interest, such as cDNA's and RNA's can be immobilized on the array or alternately used as the labeled probe mixture that is applied to the array.

In one application, an array of cDNA clones representing genes is hybridized with total cDNA from an organism to monitor gene expression for research or diagnostic purposes. Labeling total cDNA from a normal cell with one color fluorophore and total cDNA from a diseased cell with another color fluorophore and simultaneously hybridizing the two cDNA samples to the same array of cDNA clones allows for differential gene expression to be measured as the ratio of the two fluorophore intensities. This two-color experiment can be used to monitor gene expression in different tissue types, disease states, response to drugs, or response to environmental factors. & An example of this approach is illustrated in Examples 2, described with respect to Fig. 8.

By way of example and without implying a limitation of scope, such a procedure could be used to simultaneously screen many patients against all known mutations in a disease gene. This invention could be used in the form of, for example, 96 identical 0.9 cm × 2.2 cm microarrays fabricated on a single 12 cm × 18 cm

sheet of plastic-backed nitrocellulose where each microarray could contain, for example, 100 DNA fragments representing all known mutations of a given gene. The region of interest from each of the DNA samples from 96 patients could be amplified, labeled, and hybridized to the 96 individual arrays with each assay performed in 100 microliters of hybridization solution. The approximately 1 thick silicone rubber barrier elements between individual arrays prevent cross contamination of the patient samples by sealing the pores of the nitrocellulose and by acting as a physical barrier between each microarray. The solid support containing all 96 microarrays assayed with the 96 patient samples is incubated, rinsed, detected and analyzed as a single sheet of material using standard radioactive, fluorescent, or colorimetric detection means (Maniatis, et al., 1989). Previously, such a procedure would involve the handling, processing and tracking of 96 separate membranes in 96 separate sealed chambers. By processing all 96 arrays as a single sheet of material, significant time and cost savings are possible.

The assay format can be reversed where the patient or organism's DNA is immobilized as the array elements and each array is hybridized with a different mutated allele or genetic marker. The gridded solid support can also be used for parallel non-DNA ELISA assays. Furthermore, the invention allows for the use of all standard detection methods without the need to remove the shallow barrier elements to carry out the detection step.

In addition to the genetic applications listed above, arrays of whole cells, peptides, enzymes, antibodies, antigens, receptors, ligands, phospholipids, polymers, drug cogener preparations or

chemical substances can be fabricated by the means described in this invention for large scale screening assays in medical diagnostics, drug discovery, molecular biology, immunology and toxicology.

5 The multi-cell substrate aspect of the invention allows for the rapid and convenient screening of many DNA probes against many ordered arrays of DNA fragments. This eliminates the need to handle and detect many individual arrays for performing mass
10 screenings for genetic research and diagnostic applications. Numerous microarrays can be fabricated on the same solid support and each microarray reacted with a different DNA probe while the solid support is processed as a single sheet of material.

15

The following examples illustrate, but in no way are intended to limit, the present invention.

Example 1

20 Genomic-Complexity Hybridization to Micro
DNA Arrays Representing the Yeast
Saccharomyces cerevisiae Genome with
Two-Color Fluorescent Detection

The array elements were randomly amplified PCR
25 (Bohlander, et al., 1992) products using physically mapped lambda clones of *S. cerevisiae* genomic DNA templates (Riles, et al., 1993). The PCR was performed directly on the lambda phage lysates resulting in an amplification of both the 35 kb lambda vector and the
30 5-15 kb yeast insert sequences in the form of a uniform distribution of PCR product between 250-1500 base pairs in length. The PCR product was purified using Sephadex G50 gel filtration (Pharmacia, Piscataway, NJ) and concentrated by evaporation to dryness at room
35 temperature overnight. Each of the 864 amplified

lambda clones was rehydrated in 15 μ l of 3 \times SSC in preparation for spotting onto the glass.

5 The micro arrays were fabricated on microscope slides which were coated with a layer of poly-l-lysine (Sigma). The automated apparatus described in Section IV loaded 1 μ l of the concentrated lambda clone PCR product in 3 \times SSC directly from 96 well storage plates into the open capillary printing element and deposited
10 -5 nl of sample per slide at 380 micron spacing between spots, on each of 40 slides. The process was repeated for all 864 samples and 8 control spots. After the spotting operation was complete, the slides were rehydrated in a humid chamber for 2 hours, baked in a dry 80° vacuum oven for 2 hours, rinsed to remove un-
15 absorbed DNA and then treated with succinic anhydride to reduce non-specific adsorption of the labeled hybridization probe to the poly-l-lysine coated glass surface. Immediately prior to use, the immobilized DNA on the array was denatured in distilled water at 90°
20 for 2 minutes.

For the pooled chromosome experiment, the 16 chromosomes of *Saccharomyces cerevisiae* were separated in a CHEF agarose gel apparatus (Biorad, Richmond, CA). The six largest chromosomes were isolated in one gel
25 slice and the smallest 10 chromosomes in a second gel slice. The DNA was recovered using a gel extraction kit (Qiagen, Chatsworth, CA). The two chromosome pools were randomly amplified in a manner similar to that used for the target lambda clones. Following
30 amplification, 5 micrograms of each of the amplified chromosome pools were separately random-primer labeled using Klenow polymerase (Amersham, Arlington Heights, IL) with a lissamine conjugated nucleotide analog (Dupont NEN, Boston, MA) for the pool containing the
35 six largest chromosomes, and with a fluorescein

conjugated nucleotide analog (BMB) for the pool containing smallest ten chromosomes. The two pools were mixed and concentrated using an ultrafiltration device (Amicon, Danvers, MA).

5 Five micrograms of the hybridization probe consisting of both chromosome pools in 7.5 μ l of TE was denatured in a boiling water bath and then snap cooled on ice. 2.5 μ l of concentrated hybridization solution (5 \times SSC and 0.1% SDS) was added and all 10 μ l
10 transferred to the array surface, covered with a cover slip, placed in a custom-built single-slide humidity chamber and incubated at 60° for 12 hours. The slides were then rinsed at room temperature in 0.1 \times SSC and 0.1%SDS for 5 minutes, cover slipped and scanned.

15 A custom built laser fluorescent scanner was used to detect the two-color hybridization signals from the 1.8 \times 1.8 cm array at 20 micron resolution. The scanned image was gridded and analyzed using custom image analysis software. After correcting for optical
20 crosstalk between the fluorophores due to their overlapping emission spectra, the red and green hybridization values for each clone on the array were correlated to the known physical map position of the clone resulting in a computer-generated color karyotype
25 of the yeast genome.

Figure 6 shows the hybridization pattern of the two chromosome pools. A red signal indicates that the lambda clone on the array surface contains a cloned genomic DNA segment from one of the largest six yeast
30 chromosomes. A green signal indicates that the lambda clone insert comes from one of the smallest ten yeast chromosomes. Orange signals indicate repetitive sequences which cross hybridized to both chromosome pools. Control spots on the array confirm that the
35 hybridization is specific and reproducible.

The physical map locations of the genomic DNA fragments contained in each of the clones used as array elements have been previously determined by Olson and co-workers (Riles, et al.) allowing for the automatic generation of the color karyotype shown in Figure 7. The color of a chromosomal section on the karyotype corresponds to the color of the array element containing the clone from that section. The black regions of the karyotype represent false negative dark spots on the array (10%) or regions of the genome not covered by the Olson clone library (90%). Note that the largest six chromosomes are mainly red while the smallest ten chromosomes are mainly green matching the original CHEF gel isolation of the hybridization probe. Areas of the red chromosomes containing green spots and vice-versa are probably due to spurious sample tracking errors in the formation of the original library and in the amplification and spotting procedures.

The yeast genome arrays have also been probed with individual clones or pools of clones that are fluorescently labeled for physical mapping purposes. The hybridization signals of these clones to the array were translated into a position on the physical map of yeast.

25

Example 2

Total cDNA Hybridized to Micro Arrays of cDNA Clones with Two-Color Fluorescent Detection

24 clones containing cDNA inserts from the plant *Arabidopsis* were amplified using PCR. Salt was added to the purified PCR products to a final concentration of $3 \times \text{SSC}$. The cDNA clones were spotted on poly-l-lysine coated microscope slides in a manner similar to Example 1. Among the cDNA clones was a clone

representing a transcription factor HAT 4, which had previously been used to create a transgenic line of the plant *Arabidopsis*, in which this gene is present at ten times the level found in wild-type *Arabidopsis* (Schena, et al., 1992).

Total poly-A mRNA from wild type *Arabidopsis* was isolated using standard methods (Maniatis, et al., 1989) and reverse transcribed into total cDNA, using fluorescein nucleotide analog to label the cDNA product (green fluorescence). A similar procedure was performed with the transgenic line of *Arabidopsis* where the transcription factor HAT4 was inserted into the genome using standard gene transfer protocols. cDNA copies of mRNA from the transgenic plant are labeled with a lissamine nucleotide analog (red fluorescence). Two micrograms of the cDNA products from each type of plant were pooled together and hybridized to the cDNA clone array in a 10 microliter hybridization reaction in a manner similar to Example 1. Rinsing and detection of hybridization was also performed in a manner similar to Example 1. Fig. 8 show the resulting hybridization pattern of the array.

Genes equally expressed in wild type and the transgenic *Arabidopsis* appeared yellow due to equal contributions of the green and red fluorescence to the final signal. The dots are different intensities of yellow indicating various levels of gene expression. The cDNA clone representing the transcription factor HAT4, expressed in the transgenic line of *Arabidopsis* but not detectably expressed in wild type *Arabidopsis*, appears as a red dot (with the arrow pointing to it), indicating the preferential expression of the transcription factor in the red-labeled transgenic *Arabidopsis* and the relative lack of expression of the

transcription factor in the green-labeled wild type *Arabidopsis*.

An advantage of the microarray hybridization format for gene expression studies is the high partial concentration of each cDNA species achievable in the 10 microliter hybridization reaction. This high partial concentration allows for detection of rare transcripts without the need for PCR amplification of the hybridization probe which may bias the true genetic representation of each discrete cDNA species.

Gene expression studies such as these can be used for genomics research to discover which genes are expressed in which cell types, disease states, development states or environmental conditions. Gene expression studies can also be used for diagnosis of disease by empirically correlating gene expression patterns to disease states.

Example 3

Multiplexed Colorimetric Hybridization on a Gridded Solid Support

A sheet of plastic-backed nitrocellulose was gridded with barrier elements made from silicone rubber according to the description in Section IV-A. The sheet was soaked in 10 × SSC and allowed to dry. As shown in Fig. 12, 192 M13 clones each with a different yeast inserts were arrayed 400 microns apart in four quadrants of the solid support using the automated device described in Section III. The bottom left quadrant served as a negative control for hybridization while each of the other three quadrants was hybridized simultaneously with a different oligonucleotide using the open-face hybridization technology described in Section IV-A. The first two and last four elements of

each array are positive controls for the colorimetric detection step.

The oligonucleotides were labeled with fluorescein which was detected using an anti-fluorescein antibody
5 conjugated to alkaline phosphatase that precipitated an NBT/BCIP dye on the solid support (Amersham). Perfect matches between the labeled oligos and the M13 clones resulted in dark spots visible to the naked eye and detected using an optical scanner (HP ScanJet II)
10 attached to a personal computer. The hybridization patterns are different in every quadrant indicating that each oligo found several unique M13 clones from among the 192 with a perfect sequence match. Note that the open capillary printing tip leaves detectable
15 dimples on the nitrocellulose which can be used to automatically align and analyze the images.

Although the invention has been described with respect to specific embodiments and methods, it will be
20 clear that various changes and modification may be made without departing from the invention.

IT IS CLAIMED:

1. A method of forming a microarray of analyte-
assay regions on a solid support, where each region in
5 the array has a known amount of a selected, analyte-
specific reagent, said method comprising,
 (a) loading a solution of a selected analyte-
specific reagent in a reagent-dispensing device having
an elongate capillary channel (i) formed by spaced-
10 apart, coextensive elongate members, (ii) adapted to
hold a quantity of the reagent solution and (iii)
having a tip region at which aqueous solution in the
channel forms a meniscus,
 (b) tapping the tip of the dispensing device
15 against a solid support at a defined position on the
surface, with an impulse effective to break the
meniscus in the capillary channel and deposit a
selected volume of solution on the surface, and
 (c) repeating steps (a) and (b) until said array
20 is formed.
2. The method of claim 1, wherein said tapping is
carried out with an impulse effective to deposit a
selected volume in the volume range between 0.01 to 100
25 nl.
3. The method of claim 1, wherein said channel is
formed by a pair of spaced-apart tapered elements.
- 30 4. The method of claim 1, for forming a plurality
of such arrays, wherein step (b) is applied to a
selected position on each of a plurality of solid
supports at each repeat cycle proceeding step (c).

5. The method of claim 1, which further includes, after performing steps (a) and (b) at least one time, reloading the reagent-dispensing device with a new reagent solution by the steps of (i) dipping the
5 capillary channel of the device in a wash solution, (ii) removing wash solution drawn into the capillary channel, and (iii) dipping the capillary channel into the new reagent solution.

10 6. Automated apparatus for forming a microarray of analyte-assay regions on a plurality of solid supports, where each region in the array has a known amount of a selected, analyte-specific reagent, said apparatus comprising

15 (a) a holder for holding, at known positions, a plurality of planar supports,

(b) a reagent dispensing device having an open capillary channel (i) formed by spaced-apart, coextensive elongate members (ii) adapted to hold a
20 quantity of the reagent solution and (iii) having a tip region at which aqueous solution in the channel forms a meniscus,

(c) positioning means for positioning the dispensing device at a selected array position with
25 respect to a support in said holder,

(d) dispensing means for moving the device into tapping engagement against a support with a selected impulse, when the device is positioned at a defined array position with respect to that support, with an
30 impulse effective to break the meniscus of liquid in the capillary channel and deposit a selected volume of solution on the surface, and

(e) control means for controlling said positioning and dispensing means.

7. The apparatus of claim 6, wherein said dispensing means is effective to move said dispensing device against a support with an impulse effective to deposit a selected volume in the volume range between
5 0.01 to 100 nl.

8. The apparatus of claim 6, wherein said channel is formed by a pair of spaced-apart tapered elements.

10 9. The apparatus of claim 6, wherein the control means operates to (i) place the dispensing device at a loading station, (ii) move the capillary channel in the device into a selected reagent at the loading station, to load the dispensing device with the reagent, and
15 (iii) dispense the reagent at a defined array position on each of the supports on said holder.

10. The apparatus of claim 6, wherein the control device further operates, at the end of a dispensing
20 cycle, to wash the dispensing device by (i) placing the dispensing device at a washing station, (ii) moving the capillary channel in the device into a wash fluid, to load the dispensing device with the fluid, and (iii) remove the wash fluid prior to loading the dispensing
25 device with a fresh selected reagent.

11. The apparatus of claim 6, wherein said device is one of a plurality of such devices which are carried on the arm for dispensing different analyte assay
30 reagents at selected spaced array positions.

12. A substrate with a surface having a microarray of at least 10^3 distinct polynucleotide or polypeptide biopolymers per 1 cm^2 surface area, each

distinct biopolymer sample (i) being disposed at a separate, defined position in said array, (ii) having a length of at least 50 subunits, and (iii) being present in a defined amount between about 0.1 femtomole and 100 nanomoles.

13. The substrate of claim 12, wherein said surface is glass slide coated with polylysine, and said biopolymers are polynucleotides.

14. The substrate of claim 12, wherein said substrate has a water-impermeable backing, a water-permeable film formed on the backing, and a grid formed on the film, where said grid (i) is composed of intersecting water-impervious grid elements extending from said backing to positions raised above the surface of said film, and (ii) partitions the film into a plurality of water-impervious cells, where each cell contains such a biopolymer array.

15. A substrate with a surface array of sample-receiving cells, comprising
a water-impermeable backing,
a water-permeable film formed on the backing, and
a grid formed on the film, said grid being composed of intersecting water-impervious grid elements extending from said backing to positions raised above the surface of said film.

16. The substrate of claim 15, wherein the cells of the array each contain an array of biopolymers.

17. A substrate for use in detecting binding of labeled biopolymers to one or more of a plurality distinct polynucleotides, comprising

a non-porous, glass substrate,
a coating of a cationic polymer on said substrate,
and

an array of distinct polynucleotides to said
5 coating, where each biopolymer is disposed at a
separate, defined position in a surface array of
biopolymers.

18. A method of detecting differential expression
10 of each of a plurality of genes in a first cell type
with respect to expression of the same genes in a
second cell types, said method comprising

producing fluorescence-labeled cDNA's from mRNA's
isolated from the two cells types, where the cDNA's
15 from the first and second cells are labeled with first
and second different fluorescent reporters,

adding a mixture of the labeled cDNA's from the
two cell types to an array of polynucleotides
representing a plurality of known genes derived from
20 the two cell types, under conditions that result in
hybridization of the cDNA's to complementary-sequence
polynucleotides in the array; and

examining the array by fluorescence under
fluorescence excitation conditions in which (i)
25 polynucleotides in the array that are hybridized
predominantly to cDNA's derived from one of the first
and second cell types give a distinct first or second
fluorescence emission color, respectively, and (ii)
polynucleotides in the array that are hybridized to
30 substantially equal numbers of cDNA's derived from the
first and second cell types give a distinct combined
fluorescence emission color, respectively,

wherein the relative expression of known genes in
the two cell types can be determined by the observed
35 fluorescence emission color of each spot.

19. The method of claim 18, wherein the array of polynucleotides is formed on a substrate with a surface having an array of at least 10^2 distinct polynucleotide or polypeptide biopolymers in a surface area of less than about 1 cm^2 , each distinct biopolymer (i) being
5 disposed at a separate, defined position in said array, (ii) having a length of at least 50 subunits, and (iii) being present in a defined amount between about .1 femtomole and 100 nmoles.

10

20. The method of claim 19, wherein said surface is a glass slide coated with polylysine, and said biopolymers are polynucleotides non-covalently bound to said polylysine.

15

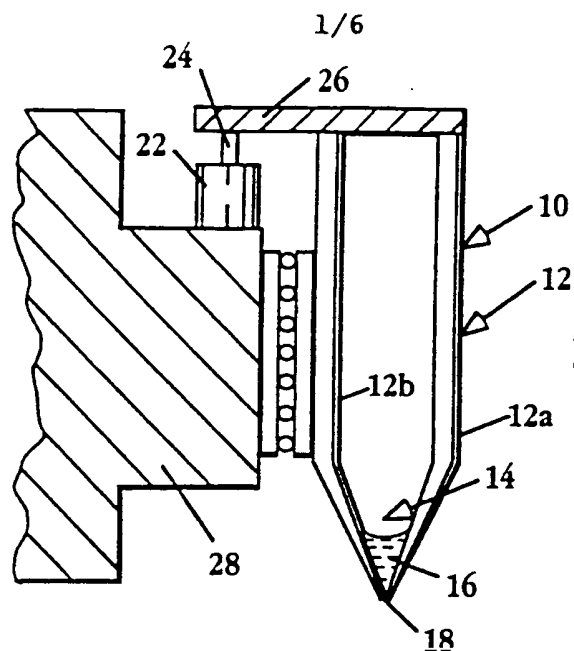


Fig. 1

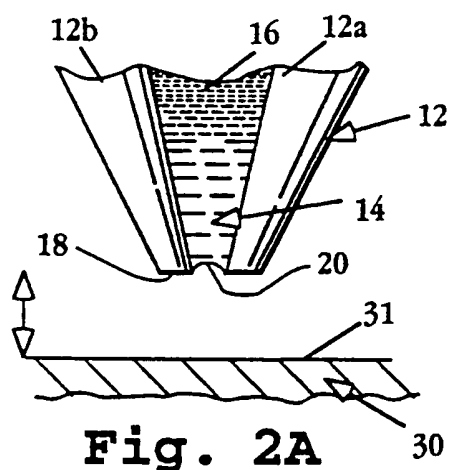


Fig. 2A

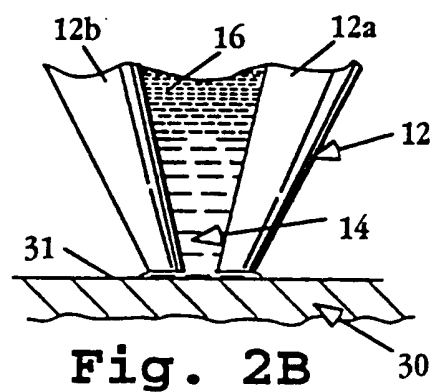


Fig. 2B

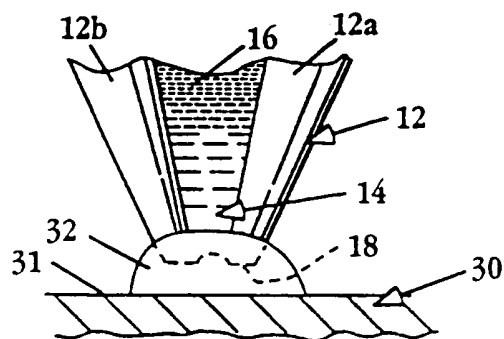


Fig. 2C

2/6

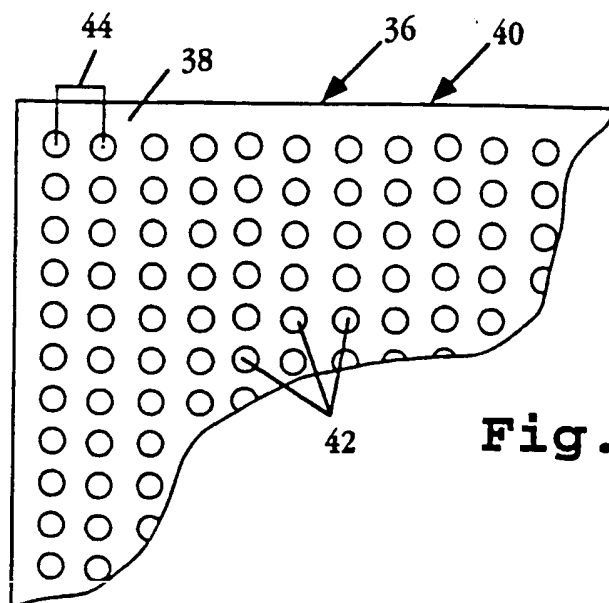


Fig. 3

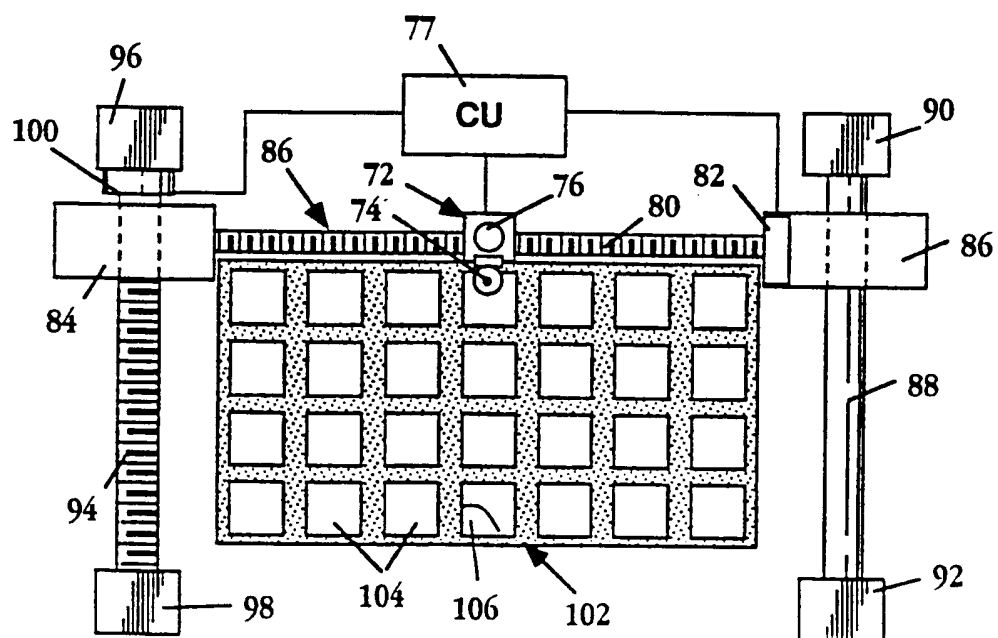


Fig. 4

3/6

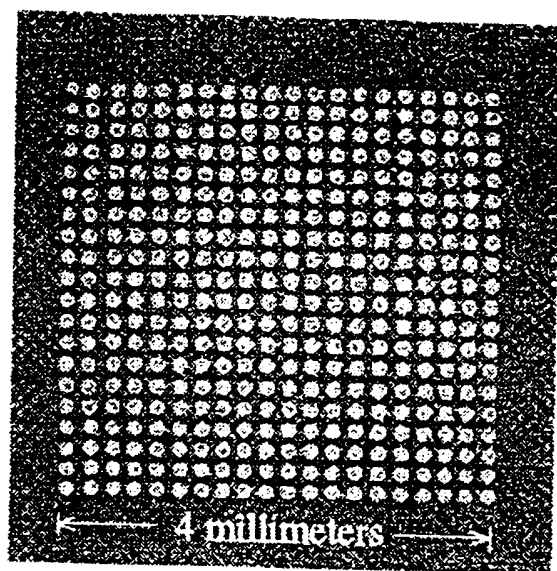


Fig. 5

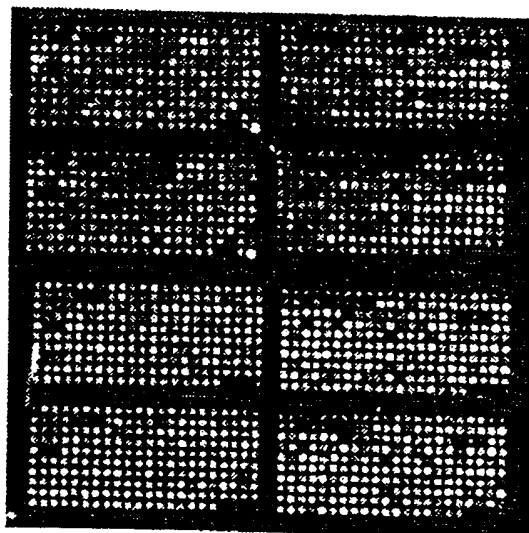


Fig. 6

4/6

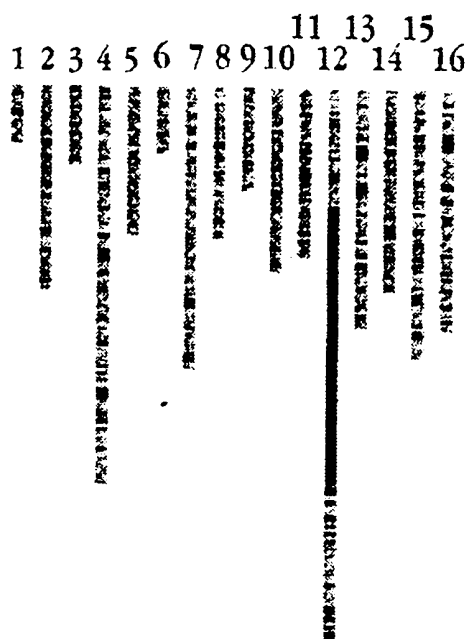


Fig. 7

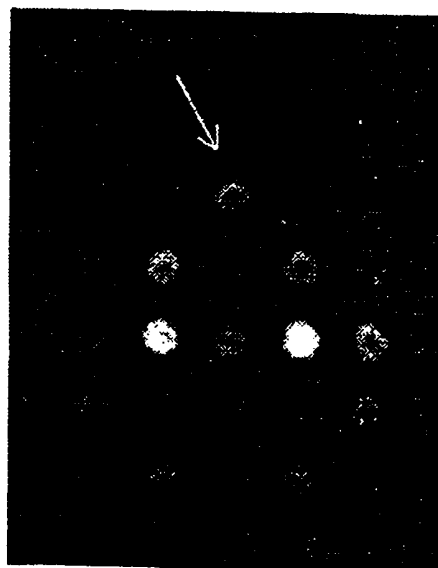


Fig. 8

5/6

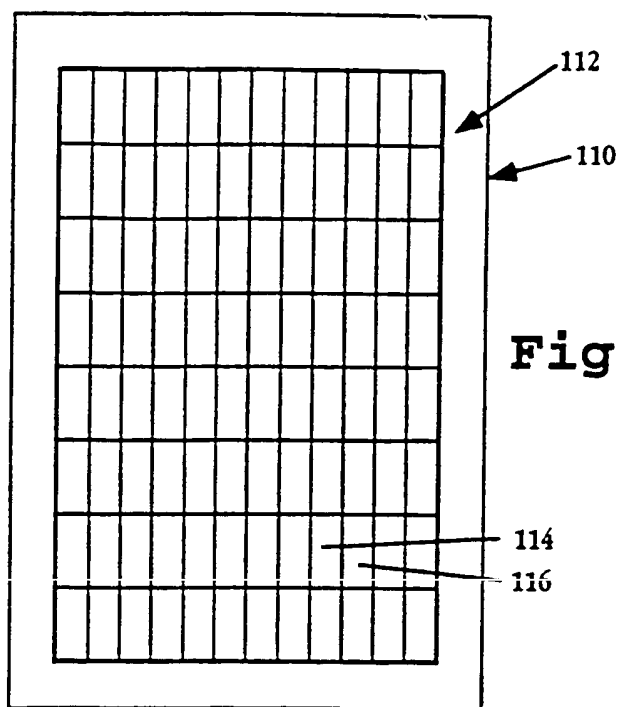


Fig. 9

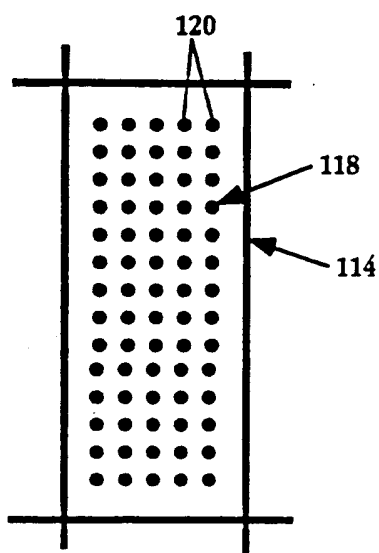
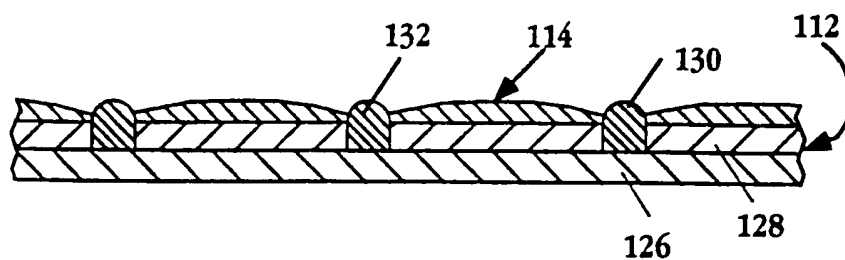
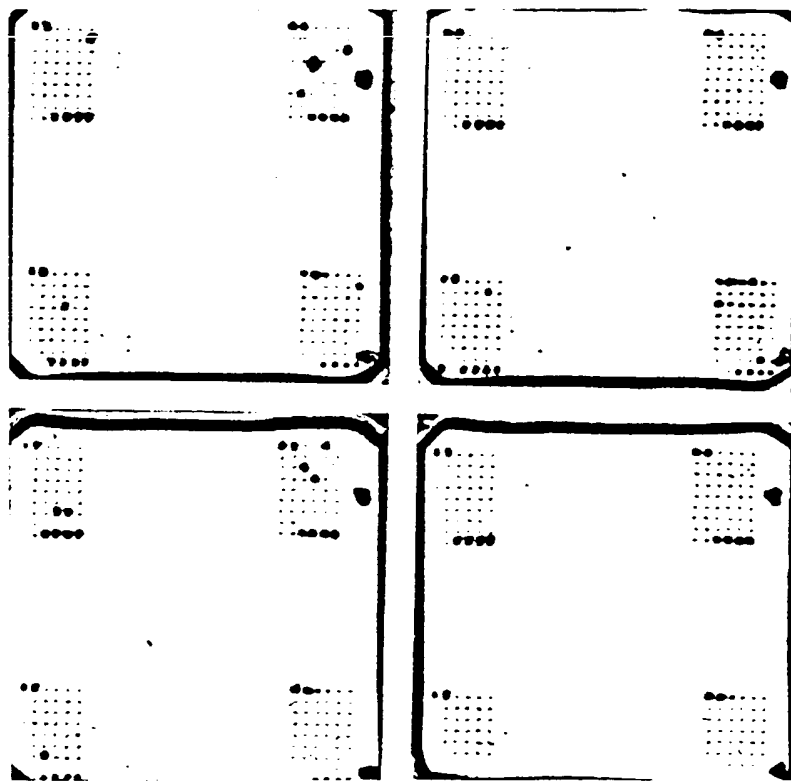


Fig. 10

6/6

**Fig. 11****Fig. 12**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US95/07659

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G01N 33/543, 33/68

US CL : 435/6; 436/518

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 422/57; 435/4.6.973; 436/518,524,527,531,805,809

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,P	US, A, 5,338,688 (DEEG ET AL) 16 August 1994, see entire document	1-17
A	US, A, 5,204,268 (MATSUMOTO) 20 April 1993, see entire document.	6-11
A	US, A, 4,071,315 (CHATEAU) 31 January 1978, see entire document.	12-17
A	US, A, 5,100,777 (CHANG) 31 March 1992, see entire document.	12-17
A	US, A, 5,200,312 (OPRANDY) 06 April 1993, see entire document.	12-17

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:	T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	
* A	document defining the general state of the art which is not considered to be of particular relevance		
* E	earlier document published on or after the international filing date	X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
* L	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
* O	document referring to an oral disclosure, use, exhibition or other means		
* P	document published prior to the international filing date but later than the priority date claimed	* &	document member of the same patent family

Date of the actual completion of the international search

15 SEPTEMBER 1995

Date of mailing of the international search report

06 OCT 1995

Name and mailing address of the International Searching Authority
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3220

Form PCT/ISA/210 (second sheet)(July 1992)*

Authorized officer

CHRISTOPHER CHIN

Telephone No. (703) 308-8196

S
SCI08
Dup

26 OCTOBER 1995

VOL. 270 • PAGES 349-500

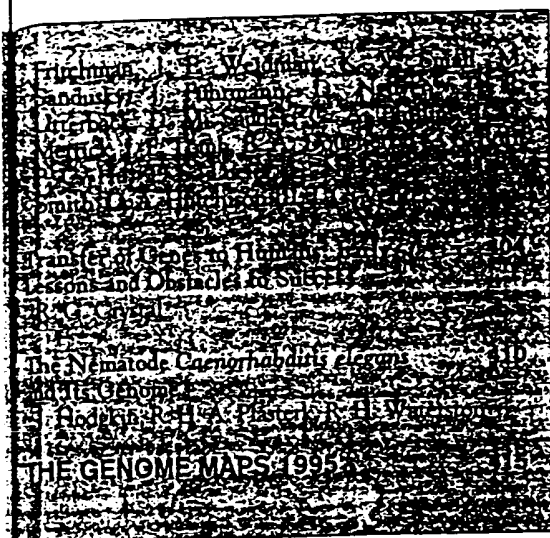
GENOME ISSUE

*****3-DIGIT 530
#025822281012\$AS 12/01/95 E 5235 74
ELDO C KOENIG
35005 W FAIRVIEW RD
OCONOMOC

COVER

The Genome Project adds a new dimension to questions on gene expression in humans and model systems. A chart on page 415 summarizes progress in the *Caenorhabditis elegans* Genome Project and indicates some ways information about sequences can be used.

News stories, Articles, Perspectives, Policy Forums, and Reports focus on technological developments, clinical applications, and ethical concerns resulting from the burgeoning of genomic information. [*C. elegans* image: F. Maduro and D. Pilgrim, University of Alberta]



REPORTS

Cosmogenic Ages for Earthquake Recurrence Intervals and Debris Flow Fan Deposition, Owens Valley, California 447
P. R. Bierman, A. R. Gillespie, M. W. Caffee

Lithoautotrophic Microbial Ecosystems in Deep Basalt Aquifers 450
T. O. Stevens and J. P. McKinley

Large Arctic Temperature Change at the Wisconsin-Holocene Glacial Transition 455
K. M. Cuffey, G. D. Clow, R. B. Alley, M. Stuiver, E. D. Waddington, R. W. Saltus

Superplasticity in Earth's Lower Mantle: Evidence from Seismic Anisotropy and Rock Physics 458
S.-i. Karato, S. Zhang, H.-R. Wenk

Large-Scale Interplanetary Magnetic Field Configuration Revealed by Solar Radio Bursts 461
M. J. Reiner, J. Fainberg, R. G. Stone

Role of Yeast Insulin-Degrading Enzyme Homologs in Pheromone Processing and Bud Site Selection 464
N. Adames, K. Blundell, M. N. Ashby, C. Boone

Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray 467
M. Schena, D. Shalon, R. W. Davis, P. O. Brown

Gene Therapy in Peripheral Blood Lymphocytes and Bone Marrow for ADA⁻ Immunodeficient Patients 470
C. Bordignon, L. D. Notarangelo, N. Nobili, G. Ferrari, G. Casorati, P. Panina, E. Mazzolari, D. Maggioni, C. Rossi, P. Servida, A. G. Ugazio, F. Mavilio

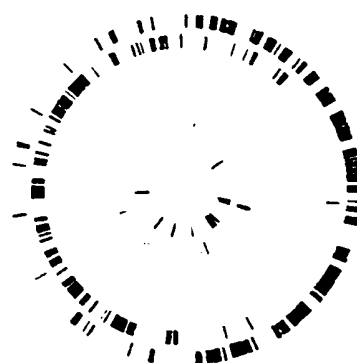
T Lymphocyte-Directed Gene Therapy for ADA⁻ SCID: Initial Trial Results After 4 Years 475
R. M. Blaese, K. W. Culver, A. D. Miller, C. S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. Chiang, P. Tolstoshev, J. J. Greenblatt, S. A. Rosenberg, H. Klein, M. Berger, C. A. Mullen, W. J. Ramsey, L. Muul, R. A. Morgan, W. F. Anderson

Physical Map and Organization of *Arabidopsis thaliana* Chromosome 4 480
R. Schmidt, J. West, K. Love, Z. Lench, C. Lister, H. Thompson, D. Bouchez, C. Dean

Serial Analysis of Gene Expression 484
V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler

TECHNICAL COMMENTS

The Radius of Gyration of an Apomyoglobin Folding Intermediate 487
D. Eliezer, P. A. Jennings, P. E. Wright, S. Doniach, K. O. Hodgson, H. Tsuruta



397
Good things in small genomes

AAAS Board of Directors ☒ Indicates accompanying feature

Francisco J. Ayala
Retiring President
1995-1996
President
1994-1995
President-elect
1995-1996

William A. Lester Jr.
Simon A. Levin
Michael J. Novacek

Anne C. Roosevelt
Alan Schlesselman
1995-1996
1995-1996
1995-1996

William T. Goode
Treasurer
Richard S. Nicholson
Executive Officer

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1333 H Street, NW, Washington, DC 20005. Second-class postage (publication No. 464460) paid at Washington, DC, and additional mailing offices. Copyright © 1995 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): \$97 (\$56 allocated to subscription). Domestic institutional subscription (51 issues): \$228 (foreign postage extra). Mexico, Caribbean (surface mail): \$50, other countries (air assist delivery): \$93. First class, airmail, student and emeritus rates on request. Canadian rates with GST available upon request. GST #1254-88122. Printed in the U.S.A.

Change of address: allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to Science, P.O. Box 1811, Danbury, CT 06813-1811. Single copy sales: \$7.00 per issue prepaid (includes surface postage; bulk rates on request). Authorization to photocopy items for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that \$3.00 per article is paid directly to CCC, 27 Congress Street, Salem, MA 01970. The identification code for Science is 0036-8075/95 \$3.00. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Axl1p sequence following Ser²⁰⁸ and occurs within the domain of Axl1p that shows homology with NDE (14). To delete the complete STE23 sequence and create the ste23Δ::URA3 mutation, polymerase chain reaction (PCR) primers (5'-TCGGAAGACCTCAT-TCTTGCTCATTTTGATATTGCTC- TGTAAGATTG-TACTGAGAGTGAC-3' and 5'-GCTACAAACAGC-GTGGACTTGAATGCCCGGACATCTTCGACTGT-GCGGTATTTCACACCG-3') were used to amplify the URA3 sequence of pRS316, and the reaction product was transformed into yeast for one-step gene replacement [R. Rothstein, *Methods Enzymol.* 194, 281 (1991)]. To create the axl1Δ::LEU2 mutation contained on p114, a 5.0-kb Sal I fragment from pAXL1 was cloned into pUC19, and an internal 4.0-kb Hpa I-Xho I fragment was replaced with a LEU2 fragment. To construct the ste23Δ::LEU2 allele (a deletion corresponding to 931 amino acids) carried on p153, a LEU2 fragment was used to replace the 2.8-kb Pml I-Ecd136 II fragment of STE23, which occurs within a 6.2-kb Hind III-Bgl II genomic fragment carried on pSP72 (Promega). To create YEPMAF1, a 1.6-kb Bam HI fragment containing MAF1, from pKK16 [K. Kuchler, R. E. Sterne, J. Thorne, *EMBO J.* 8, 3973 (1989)], was ligated into the Bam HI site of YEP351 [J. E. Hill, A. M. Myers, T. J. Koerner, A. Tzagoloff, *Yeast* 2, 163 (1986)].

24. J. Chant and I. Herskowitz, *Cell* 65, 1203 (1991).
25. B. W. Matthews, *Acc. Chem. Res.* 21, 333 (1988).
26. K. Kuchler, H. G. Dohman, J. Thorne, *J. Cell Biol.* 120, 1203 (1993); R. Koling and C. P. Hollenberg, *EMBO J.* 13, 3261 (1994); C. Berkower, D. Loayza, S. Michaelis, *Mol. Biol. Cell* 5, 1185 (1994).
27. A. Bender and J. R. Pringle, *Proc. Natl. Acad. Sci. U.S.A.* 86, 9976 (1989); J. Chant, K. Corrado, J. R. Pringle, I. Herskowitz, *Cell* 65, 1213 (1991); S. Powers, E. Gonzales, T. Christensen, J. Cubert, D. Broek, *ibid.*, p. 1225; H. O. Park, J. Chant, I. Herskowitz, *Nature* 365, 269 (1993); J. Chant, *Trends Genet.* 10, 328 (1994); J. R. Pringle, *J. Cell Biol.* 129, 751 (1995); J. Chant, M. Mischke, E. Mitchell, I. Herskowitz, J. R. Pringle, *ibid.*, p. 767.
28. G. F. Sprague Jr., *Methods. Enzymol.* 194, 77 (1991).
29. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
30. A W303 1A derivative, SY2625 (MATa ura3-1 leu2-3, 112 trp1-1 ade2-1 can1-100 sst1Δ mfa2Δ::FUS1-lacZ his3Δ::FUS1-HIS3), was the parent strain for the mutant strains. SY2625 derivatives for the mating assays, secreted pheromone assays, and the pulse-chase experiments included the following strains: Y49 (ste22-1), Y115 (mfa1Δ::LEU2), Y142 (axl1::URA3), Y173 (axl1Δ::LEU2), Y220 (axl1::URA3 ste23Δ::URA3), Y221 (ste23Δ::URA3), Y231 (axl1Δ::LEU2 ste23Δ::LEU2), and Y233 (ste23Δ::LEU2). MATa derivatives of SY2625 included the following strains: Y199 (SY2625 made MATa), Y278 (ste22-1), Y195 (mfa1Δ::LEU2), Y196 (axl1Δ::LEU2), and Y197 (axl1::URA3). The EG123 (MATa leu2 ura3 trp1 can1 his4) genetic background was used to create a set of strains for analysis of bud site selection. EG123 derivatives included the following strains: Y175 (axl1Δ::LEU2), Y223 (axl1::URA3), Y234 (ste23Δ::LEU2), and Y272 (axl1Δ::LEU2 ste23Δ::LEU2). MATa derivatives of EG123 included the following strains: Y214 (EG123 made MATa) and Y293 (axl1Δ::LEU2). All strains were generated by means of standard genetic or molecular methods involving the appropriate constructs (23). In particular, the axl1 ste23 double mutant strains were created by crossing of the appropriate MATa ste23 and MATa axl1 mutants, followed by sporulation of the resultant diploid and isolation of the double mutant from nonparental tetrads. Gene disruptions were created with either the homologous recombination (HCR) method or a YEP352 (J. E. Hill, A. M. Myers, T. J. Koerner, A. Tzagoloff, *Yeast* 2, 163 (1986)) plasmid containing a 5.5-kb Sal I fragment of pAXL1. p151 was derived from p129 by insertion of a linker at the Bgl II site within AXL1, which led to an in-frame insertion of the hemagglutinin (HA) epitope (DQYPYDVPDYA) (29) between amino acids 854 and 855 of the AXL1 prod-

uct. pC225 is a KS+ (Stratagene) plasmid containing a 0.5-kb Bam HI-Sst I fragment from pAXL1. Substitution mutations of the proposed active site of Axl1p were created with the use of pC225 and site-specific mutagenesis involving appropriate synthetic oligonucleotides (axl1-H68A, 5'-GTGCTCACAAGCGCT-GCCAAACCGGC-3'; axl1-E71A, 5'-AAGAATCAT-GTGGCACAAGGTGGCG-3'; and axl1-E71D, 5'-AAGAATCATGTGATCACAAGGTGGCG-3'). The mutations were confirmed by sequence analysis. After mutagenesis, the 0.4-kb Bam HI-Msc I fragment from the mutagenized pC225 plasmids was transferred into pAXL1 to create a set of pRS316 plasmids carrying different AXL1 alleles, p124 (axl1-H68A), p130 (axl1-E71A), and p132 (axl1-E71D). Similarly, a set of HA-tagged alleles carried on YEP352 were created after replacement of the p151 Bam HI-Msc I fragment, to generate p161 (axl1-E71A), p162 (axl1-

H68A), and p163 (axl1-E71D).

32. We thank J. Becker and S. Michaelis for providing a-factor antibodies; S. Michaelis for discussing unpublished results and helping with the pulse-chase experiments; J. Brown, J. Chant, and S. Sanders for their input concerning bud site selection experiments; M. Raymond, F. Tamanoi, and M. Whiteway for plasmids; M. Marra for providing the STE23 genomic fragment; and H. Bussey, J. Brown, N. Davis, T. Favero, C. de Hoog, and S. Kim for comments on the manuscript. Supported by a grant to C.B. from the Natural Sciences and Engineering Research Council of Canada. Support for M.N.A. was from a California Tobacco-Related Disease Research Program postdoctoral fellowship (4FT-0083).

22 June 1995; accepted 21 August 1995

Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray

Mark Schena,* Dari Shalon,†† Ronald W. Davis, Patrick O. Brown‡

A high-capacity system was developed to monitor the expression of many genes in parallel. Microarrays prepared by high-speed robotic printing of complementary DNAs on glass were used for quantitative expression measurements of the corresponding genes. Because of the small format and high density of the arrays, hybridization volumes of 2 microliters could be used that enabled detection of rare transcripts in probe mixtures derived from 2 micrograms of total cellular messenger RNA. Differential expression measurements of 45 *Arabidopsis* genes were made by means of simultaneous, two-color fluorescence hybridization.

The temporal, developmental, topographical, histological, and physiological patterns in which a gene is expressed provide clues to its biological role. The large and expanding database of complementary DNA (cDNA) sequences from many organisms (1) presents the opportunity of defining these patterns at the level of the whole genome.

For these studies, we used the small flowering plant *Arabidopsis thaliana* as a model organism. *Arabidopsis* possesses many advantages for gene expression analysis, including the fact that it has the smallest genome of any higher eukaryote examined to date (2). Forty-five cloned *Arabidopsis* cDNAs (Table 1), including 14 complete sequences and 31 expressed sequence tags (ESTs), were used as gene-specific targets. We obtained the ESTs by selecting cDNA clones at random from an *Arabidopsis* cDNA library. Sequence analysis revealed that 28 of the 31 ESTs matched sequences

in the database (Table 1). Three additional cDNAs from other organisms served as controls in the experiments.

The 48 cDNAs, averaging ~1.0 kb, were amplified with the polymerase chain reaction (PCR) and deposited into individual wells of a 96-well microtiter plate. Each sample was duplicated in two adjacent wells to allow the reproducibility of the arraying and hybridization process to be tested. Samples from the microtiter plate were printed onto glass microscope slides in an area measuring 3.5 mm by 5.5 mm with the use of a high-speed arraying machine (3). The arrays were processed by chemical and heat treatment to attach the DNA sequences to the glass surface and denature them (3). Three arrays, printed in a single lot, were used for the experiments here. A single microtiter plate of PCR products provides sufficient material to print at least 500 arrays.

Fluorescent probes were prepared from total *Arabidopsis* mRNA (4) by a single round of reverse transcription (5). The *Arabidopsis* mRNA was supplemented with human acetylcholine receptor (AChR) mRNA at a dilution of 1:10,000 (w/w) before cDNA synthesis, to provide an internal standard for calibration (5). The resulting fluorescently labeled cDNA mixture was hybridized to an array at high stringency (6) and scanned

M. Schena and R. W. Davis, Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305, USA.

D. Shalon, Department of Biochemistry, Howard Hughes Medical Institute, Beckman Center, Stanford University Medical Center, Stanford, CA 94305, USA.

*These authors contributed equally to this work.

†Present address: Syntex, Palo Alto, CA 94303, USA.

‡To whom correspondence should be addressed. E-mail: pbrown@crgm.stanford.edu

with a laser (3). A high-sensitivity scan gave signals that saturated the detector at nearly all of the *Arabidopsis* target sites (Fig. 1A). Calibration relative to the AChR mRNA standard (Fig. 1A) established a sensitivity limit of $\sim 1:50,000$. No detectable hybridization was observed to either the rat glucocorticoid receptor (Fig. 1A) or the yeast TRP4 (Fig. 1A) targets even at the highest scanning sensitivity. A moderate-sensitivity scan

of the same array allowed linear detection of the more abundant transcripts (Fig. 1B). Quantitation of both scans revealed a range of expression levels spanning three orders of magnitude for the 45 genes tested (Table 2). RNA blots (7) for several genes (Fig. 2) corroborated the expression levels measured with the microarray to within a factor of 5 (Table 2).

Differential gene expression was investi-

gated with a simultaneous, two-color hybridization scheme, which served to minimize experimental variation inherent in the comparison of independent hybridizations. Fluorescent probes were prepared from two mRNA sources with the use of reverse transcriptase in the presence of fluorescein- and lissamine-labeled nucleotide analogs, respectively (5). The two probes were then mixed together in equal proportions, hybridized to a single array, and scanned separately for fluorescein and lissamine emission after independent excitation of the two fluorophores (3).

To test whether overexpression of a single gene could be detected in a pool of total *Arabidopsis* mRNA, we used a microarray to analyze a transgenic line overexpressing the single transcription factor HAT4 (8). Fluorescent probes representing mRNA from wild-type and HAT4-transgenic plants were labeled with fluorescein and lissamine, respectively; the two probes were then mixed and hybridized to a single array. An intense hybridization signal was observed at the position of the HAT4 cDNA in the lissamine-specific scan (Fig. 1D), but not in the fluorescein-specific scan of the same array (Fig. 1C). Calibration with AChR mRNA added to the fluorescein and lissamine cDNA synthesis reactions at dilutions of 1:10,000 (Fig. 1C) and 1:100 (Fig. 1D), respectively, revealed a 50-fold elevation of HAT4 mRNA in the transgenic line relative to its abundance in wild-type plants (Table 2). This magnitude of HAT4 overexpression matched that inferred from the Northern (RNA) analysis within a factor of 2 (Fig. 2 and Table 2). Expression of all the other genes monitored on the array differed by less than a factor of 5 between HAT4-transgenic and wild-type plants (Fig. 1, C

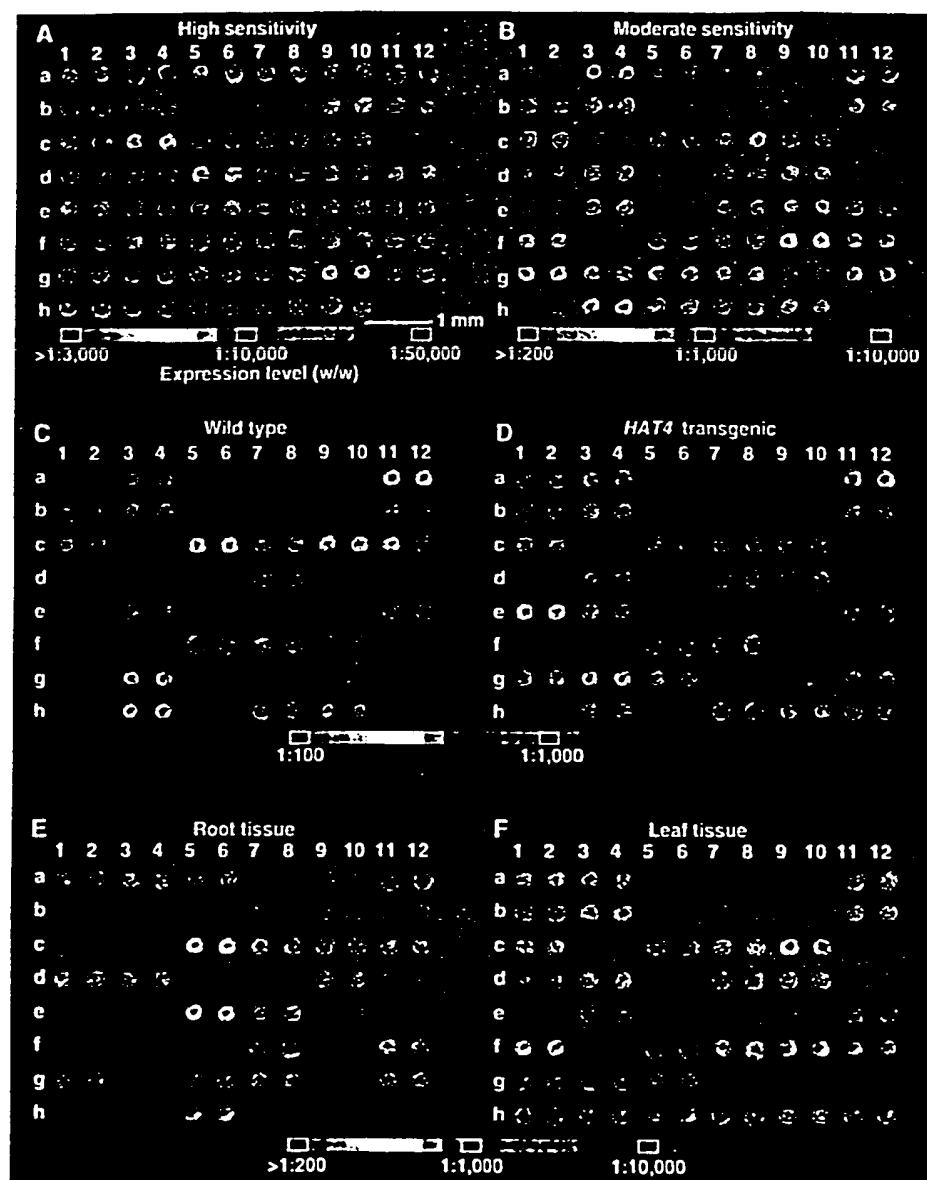


Fig. 1. Gene expression monitored with the use of cDNA microarrays. Fluorescent scans represented in pseudocolor correspond to hybridization intensities. Color bars were calibrated from the signal obtained with the use of known concentrations of human AChR mRNA in independent experiments. Numbers and letters on the axes mark the position of each cDNA. (A) High-sensitivity fluorescein scan after hybridization with fluorescein-labeled cDNA derived from wild-type plants. (B) Same array as in (A) but scanned at moderate sensitivity. (C and D) A single array was probed with a 1:1 mixture of fluorescein-labeled cDNA from wild-type plants and lissamine-labeled cDNA from HAT4-transgenic plants. The single array was then scanned successively to detect the fluorescein fluorescence corresponding to mRNA from wild-type plants (C) and the lissamine fluorescence corresponding to mRNA from HAT4-transgenic plants (D). (E and F) A single array was probed with a 1:1 mixture of fluorescein-labeled cDNA from root tissue and lissamine-labeled cDNA from leaf tissue. The single array was then scanned successively to detect the fluorescein fluorescence corresponding to mRNAs expressed in roots (E) and the lissamine fluorescence corresponding to mRNAs expressed in leaves (F).

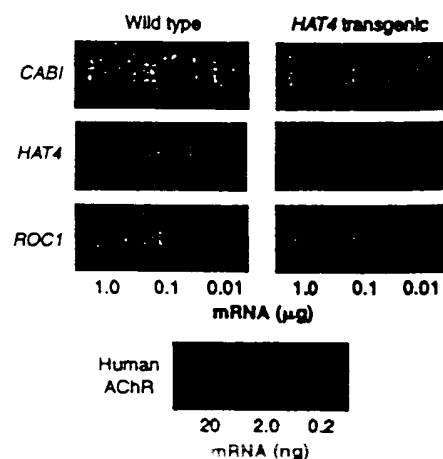


Fig. 2. Gene expression monitored with RNA Northern blot analysis. Designated amounts of mRNA from wild-type and HAT4-transgenic plants were spotted onto nylon membranes and probed with the cDNAs indicated. Purified human AChR mRNA was used for calibration.

and D, and Table 2). Hybridization of fluorescein-labeled glucocorticoid receptor cDNA (Fig. 1C) and lissamine-labeled TRP4 cDNA (Fig. 1D) verified the presence of the negative control targets and the lack of optical cross talk between the two fluorophores.

To explore a more complex alteration in expression patterns, we performed a second two-color hybridization experiment with fluorescein- and lissamine-labeled probes prepared from root and leaf mRNA, respectively. The scanning sensitivities for the two fluorophores were normalized by matching the signals resulting from AChR

mRNA, which was added to both cDNA synthesis reactions at a dilution of 1:1000 (Fig. 1, E and F). A comparison of the scans revealed widespread differences in gene expression between root and leaf tissue (Fig. 1, E and F). The mRNA from the light-regulated CAB1 gene was ~500-fold more abundant in leaf (Fig. 1F) than in root tissue (Fig. 1E). The expression of 26 other genes differed between root and leaf tissue by more than a factor of 5 (Fig. 1, E and F).

The HAT4-transgenic line we examined has elongated hypocotyls, early flowering, poor germination, and altered pigmentation (8). Although changes in expression were

observed for HAT4, large changes in expression were not observed for any of the other 44 genes we examined. This was somewhat surprising, particularly because comparative analysis of leaf and root tissue identified 27 differentially expressed genes. Analysis of an expanded set of genes may be required to identify genes whose expression changes upon HAT4 overexpression; alternatively, a comparison of mRNA populations from specific tissues of wild-type and HAT4-transgenic plants may allow identification of downstream genes.

At the current density of robotic printing, it is feasible to scale up the fabrication process to produce arrays containing 20,000 cDNA targets. At this density, a single array would be sufficient to provide gene-specific targets encompassing nearly the entire repertoire of expressed genes in the *Arabidopsis* genome (2). The availability of 20,274 ESTs from *Arabidopsis* (1, 9) would provide a rich source of templates for such studies.

The estimated 100,000 genes in the human genome (10) exceeds the number of *Arabidopsis* genes by a factor of 5 (2). This modest increase in complexity suggests that similar cDNA microarrays, prepared from the rapidly growing repertoire of human ESTs (1), could be used to determine the expression patterns of tens of thousands of human genes in diverse cell types. Coupling an amplification strategy to the reverse transcription reaction (11) could make it feasible to monitor expression even in minute tissue samples. A wide variety of acute and chronic physiological and pathological conditions might lead to characteristic changes in the patterns of gene expression in peripheral blood cells or other easily sampled tissues. In concert with cDNA microarrays for monitoring complex expression patterns, these tissues might therefore serve as sensitive *in vivo* sensors for clinical diagnosis. Microarrays of cDNAs could thus provide a useful link between human gene sequences and clinical medicine.

Table 2. Gene expression monitoring by microarray and RNA blot analyses; tg, HAT4-transgenic. See Table 1 for additional gene information. Expression levels (w/w) were calibrated with the use of known amounts of human AChR mRNA. Values for the microarray were determined from microarray scans (Fig. 1); values for the RNA blot were determined from RNA blots (Fig. 2).

Gene	Expression level (w/w)	
	Microarray	RNA blot
AChR	1	1
ACT1	1	50
ACT2	1	150
HAT4	1:8300	1:6300
HAT4 (tg)	1:150	1:210
ROC1	1:1200	1:1800
ROC1 (tg)	1:260	1:1300

Table 1. Sequences contained on the cDNA microarray. Shown is the position, the known or putative function, and the accession number of each cDNA in the microarray (Fig. 1). All but three of the ESTs used in this study matched a sequence in the database. NADH, reduced form of nicotinamide adenine dinucleotide; ATPase, adenosine triphosphatase; GTP, guanosine triphosphate.

Position	cDNA	Function	Accession number
a1, 2	AChR	Human AChR	-
a3, 4	EST3	Actin	H36236
a5, 6	EST6	NADH dehydrogenase	Z27010
a7, 8	AAC1	Actin 1	M20016
a9, 10	EST12	Unknown	U36594†
a11, 12	EST13	Actin	T45783
b1, 2	CAB1	Chlorophyll a/b binding	M85150
b3, 4	EST17	Phosphoglycerate kinase	T44490
b5, 6	GA4	Gibberellic acid biosynthesis	L37126
b7, 8	EST19	Unknown	U36595†
b9, 10	GBF-1	G-box binding factor 1	X63894
b11, 12	EST23	Elongation factor	X52256
c1, 2	EST29	Aldolase	T04477
c3, 4	GBF-2	G-box binding factor 2	X63895
c5, 6	EST34	Chloroplast protease	R87034
c7, 8	EST35	Unknown	T14152
c9, 10	EST41	Catalase	T22720
c11, 12	rGR	Rat glucocorticoid receptor	M14053
d1, 2	EST42	Unknown	U36596†
d3, 4	EST45	ATPase	J04185
d5, 6	HAT1	Homeobox-leucine zipper 1	U09332
d7, 8	EST46	Light harvesting complex	T04063
d9, 10	EST49	Unknown	T76267
d11, 12	HAT2	Homeobox-leucine zipper 2	U09335
e1, 2	HAT4	Homeobox-leucine zipper 4	M90394
e3, 4	EST50	Phosphoribulokinase	T04344
e5, 6	HAT5	Homeobox-leucine zipper 5	M90416
e7, 8	EST51	Unknown	Z33675
e9, 10	HAT22	Homeobox-leucine zipper 22	U09336
e11, 12	EST52	Oxygen evolving	T21749
f1, 2	EST59	Unknown	Z34607
f3, 4	KNAT1	Knotted-like homeobox 1	U14174
f5, 6	EST60	RuBisCO small subunit	X14564
f7, 8	EST69	Translation elongation factor	T42799
f9, 10	PPH1	Protein phosphatase 1	U34803
f11, 12	EST70	Unknown	T44621
g1, 2	EST75	Chloroplast protease	T43698
g3, 4	EST78	Unknown	R65481
g5, 6	ROC1	Cyclophilin	L14844
g7, 8	EST82	GTP binding	X59152
g9, 10	EST83	Unknown	Z33795
g11, 12	EST84	Unknown	T45278
h1, 2	EST9	Unknown	T3831
h3, 4	EST10	Unknown	T4481†
h5, 6	EST11	Chlorophyll	X9041†
h7, 8	EST12	Light harvesting complex	Z18205
h9, 10	EST103	Light harvesting complex	X03909
h11, 12	TRP4	Yeast tryptophan biosynthesis	X04273

*Proprietary sequence of Stratagene (La Jolla, California).

†No match in the database; novel EST.

REFERENCES AND NOTES

The current EST database (dbEST release 091495) from the National Center for Biotechnology Information (Bethesda, MD) contains a total of 322,225 entries, including 255,645 from the human genome and 21,044 from *Arabidopsis*. Access is available via the World Wide Web (<http://www.ncbi.nlm.nih.gov>). E. M. Meyerowitz and R. E. Pruitt, *Science* 229, 1214 (1985); R. E. Pruitt and E. M. Meyerowitz, *J. Mol. Biol.* 187, 169 (1986); I. Hwang et al., *Plant J.* 1, 367 (1991); P. Jarvis et al., *Plant Mol. Biol.* 24, 685 (1994); L. Le Guen et al., *Mol. Gen. Genet.* 245, 390 (1994). D. Shalon, thesis, Stanford University (1995); and P. O. Brown, in preparation. Microarrays were fabricated on poly-L-lysine-coated microscope slides (Sigma) with a custom-built arraying machine fitted with one printing tip. The tip loaded 1 μ l of PCR product (0.5 mg/ml) from 96-well microtiter plates and deposited \sim 0.005 μ l per slide on 40 slides at a spacing of 500 μ m. The printed slides were rehydrated for 2 hours in a humid chamber, snap-dried at 100°C for 1 min, rinsed in 0.1% SDS, and treated with 0.05% succinic anhydride prepared in buffer consisting of 50% 1-methyl-2-pyrrolidinone and 50% boric acid. The cDNA on the slides was denatured in distilled water for 2 min at 90°C immediately before use. Microarrays were scanned with a laser fluorescent scanner that contained a computer-controlled XY stage and a microscope objective. A mixed gas, multiline laser allowed sequential excitation of the two fluorophores. Emitted light was split according to wavelength and detected with two photomultiplier tubes. Signals were read into a PC with the use of a 12-bit analog-to-digital board. Additional details of microarray fabrication and use may be obtained by means of e-mail (pbrown@cimgm.stanford.edu).

F. M. Ausubel et al., Eds., *Current Protocols in Molecular Biology* (Greene & Wiley Interscience, New York, 1994), pp. 4.3.1–4.3.4.

Polyadenylated [poly(A)⁺] mRNA was prepared from total RNA with the use of Oligotex-dT resin (Qiagen). Reverse transcription (RT) reactions were carried out with a StrataScript RT-PCR kit (Stratagene) modified as follows: 50- μ l reactions contained 0.1 μ g/ μ l of *Arabidopsis* mRNA, 0.1 ng/ μ l of human AChR mRNA, 0.05 μ g/ μ l of oligo(dT) (21-mer), 1 \times first strand buffer, 0.03 U/ μ l of ribonuclease block, 500 μ M deoxyadenosine triphosphate (dATP), 500 μ M deoxyguanosine triphosphate, 500 μ M dTTP, 40 μ M deoxycytosine triphosphate (dCTP), 40 μ M fluorescein-12-dCTP (or lissamine-5-dCTP), and 0.03 U/ μ l of StrataScript reverse transcriptase. Reactions were incubated for 60 min at 37°C, precipitated with ethanol, and resuspended in 10 μ l of TE (10 mM Tris-HCl and 1 mM EDTA, pH 8.0). Samples were then heated for 3 min at 94°C and chilled on ice. The RNA was degraded by adding 0.25 μ l of 10 N NaOH followed by a 10-min incubation at 37°C. The samples were neutralized by addition of 2.5 μ l of 1 M Tris-HCl (pH 8.0) and 0.25 μ l of 10 N HCl and precipitated with ethanol. Pellets were washed with 70% ethanol, dried to completion in a speedvac, resuspended in 10 μ l of H₂O, and reduced to 3.0 μ l in a speedvac. Fluorescent nucleotide analogs were obtained from New England Nuclear (DuPont).

Hybridization reactions contained 1.0 μ l of fluorescent DNA synthesis product (5) and 1.0 μ l of hybridization buffer [10 \times saline sodium citrate (SSC) and 0.2% DS]. The 2.0- μ l probe mixtures were aliquoted onto the microarray surface and covered with cover slips (2 mm round). Arrays were transferred to a hybridization chamber (3) and incubated for 18 hours at 5°C. Arrays were washed for 5 min at room temperature (25°C) in low-stringency wash buffer (1 \times SSC + 0.1% SDS), then for 10 min at room temperature in high-stringency wash buffer (0.1 \times SSC and 0.1% SDS). Arrays were scanned in 0.1 \times SSC with the use of a line-scanning laser scanning device (3).

Microarrays were fixed with glutaraldehyde, membranes hybridized with digoxigenin-labeled probes, and detected with the use of a StrataLinker (Boehringer-Mannheim). Probes were prepared by random priming with the use of a Prime-It II kit (Stratagene) in the presence of [³²P]dATP. Hybridizations were carried out according to the instructions of the manu-

- facturer. Quantitation was performed on a PhosphorImager (Molecular Dynamics).
8. M. Schena and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* 89, 3894 (1992); M. Schena, A. M. Lloyd, R. W. Davis, *Genes Dev.* 7, 367 (1993); M. Schena and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* 91, 8393 (1994).
9. H. Hofte et al., *Plant J.* 4, 1051 (1993); T. Newman et al., *Plant Physiol.* 106, 1241 (1994).
10. N. E. Morton, *Proc. Natl. Acad. Sci. U.S.A.* 88, 7474 (1991); E. D. Green and R. H. Waterston, *J. Am. Med. Assoc.* 266, 1966 (1991); C. Betanhe-Charrelot, *Cell* 70, 1059 (1992); D. R. Cox et al., *Science* 265, 2031 (1994).
11. E. S. Kawasaki et al., *Proc. Natl. Acad. Sci. U.S.A.* 85, 5698 (1988).

12. The laser fluorescent scanner was designed and fabricated in collaboration with S. Smith of Stanford University. Scanner and analysis software was developed by R. X. Xia. The succinic anhydride reaction was suggested by J. Mulligan and J. Van Ness of Darwin Molecular Corporation. Thanks to S. Theologis, C. Somerville, K. Yamamoto, and members of the laboratories of R. W. C. and P. O. B. for critical comments. Supported by the Howard Hughes Medical Institute and by grants from NIH (R21HG00450) (P.O.B.) and R37AG00198 (R.W.D.) and from NSF (MCB9106011) (R.W.D.) and by an NSF graduate fellowship (D.S.). P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

11 August 1995; accepted 22 September 1995

Gene Therapy in Peripheral Blood Lymphocytes and Bone Marrow for ADA⁻ Immunodeficient Patients

Claudio Bordignon,* Luigi D. Notarangelo, Nadia Nobili, Giuliana Ferrari, Giulia Casorati, Paola Panina, Evelina Mazzolari, Daniela Maggioni, Claudia Rossi, Paolo Servida, Alberto G. Ugazio, Fulvio Mavilio

Adenosine deaminase (ADA) deficiency results in severe combined immunodeficiency, the first genetic disorder treated by gene therapy. Two different retroviral vectors were used to transfer ex vivo the human ADA minigene into bone marrow cells and peripheral blood lymphocytes from two patients undergoing exogenous enzyme replacement therapy. After 2 years of treatment, long-term survival of T and B lymphocytes, marrow cells, and granulocytes expressing the transferred ADA gene was demonstrated and resulted in normalization of the immune repertoire and restoration of cellular and humoral immunity. After discontinuation of treatment, T lymphocytes, derived from transduced peripheral blood lymphocytes, were progressively replaced by marrow-derived T cells in both patients. These results indicate successful gene transfer into long-lasting progenitor cells, producing a functional multilineage progeny.

Severe combined immunodeficiency associated with inherited deficiency of ADA (1) is usually fatal unless affected children are kept in protective isolation or the immune system is reconstituted by bone marrow transplantation from a human leukocyte antigen (HLA)-identical sibling donor (2). This is the therapy of choice, although it is available only for a minority of patients. In recent years, other forms of therapy have been developed, including transplants from haploidentical donors (3, 4), exogenous enzyme replacement (5), and somatic-cell gene therapy (6–9).

We previously reported a preclinical model in which ADA gene transfer and expression

successfully restored immune functions in human ADA-deficient (ADA⁻) peripheral blood lymphocytes (PBLs) in immunodeficient mice in vivo (10, 11). On the basis of these preclinical results, the clinical application of gene therapy for the treatment of ADA⁻ SCID (severe combined immunodeficiency disease) patients who previously failed exogenous enzyme replacement therapy was approved by our Institutional Ethical Committees and by the Italian National Committee for Bioethics (12). In addition to evaluating the safety and efficacy of the gene therapy procedure, the aim of the study was to define the relative role of PBLs and hematopoietic stem cells in the long-term reconstitution of immune functions after retroviral vector-mediated ADA gene transfer. For this purpose, two structurally identical vectors expressing the human ADA complementary DNA

with alternative restriction sites in a nonfunctional region of the viral long-terminal repeat (LTR), were used to transduce PBLs and bone marrow (BM) cells independently. This procedure allowed identification of the origin of

C. Bordignon, N. Nobili, G. Ferrari, D. Maggioni, C. Rossi, P. Servida, F. Mavilio, Telethon Gene Therapy Program for Genetic Diseases, DIBIT, Istituto Scientifico H. S. Raffaele, Milan, Italy.

L. D. Notarangelo, F. Mavilio, Telethon Gene Therapy Program for Genetic Diseases, DIBIT, Istituto Scientifico H. S. Raffaele, Milan, Italy.

G. Casorati, Unità di Immunochimica, DIBIT, Istituto Scientifico H. S. Raffaele, Milan, Italy.

P. Panina, Roche Milano Ricerche, Milan, Italy.

*To whom correspondence should be addressed.

nature *genetics*

volume 14 no. 4

december 1996

**DNA chips,
diagnostics
and genomics**

**Rieger
syndrome**

**QTLs and
epistasis**

**A BRCA1-
binding
protein**



LIBRARY OF MED BOX 1102
FIBER CG 11TH FL
GUSTAV LEVY PL
YOR 10029-6504

00/003/0009



Nature Genetics

Editor
Kevin Davies

Assistant Editors
Laure Goodman
Bette Phimister

Production Editor
J. Stuart Griffith

Assistant Production Editor
Ken Krattenmaker

Editorial Assistant
Janelle Bolden

Washington Bureau Chief
Barbara J. Culliton

Editorial Office
545 National Press Building
Washington DC 20045
Tel: (202) 626-2513
Fax: (202) 626-0970
email: natgen@naturedc.com

WWW: genetics.nature.com



Cover art: Ken Krattenmaker

editorial To affinity . . . and beyond! **edit** 367

news & views Who's afraid of epistasis? 371

Wayne N Frankel & Nicholas J Schork **ncv**

Meiotic nondisjunction does the two-step 374

Terry Orr-Weaver **ncv**

Flood warning — resistance genes unleashed 376

Richard Michelmore **ncv**

correspondence Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence 380

A H Paterson, T-H Lan, K P Reischmann, C Chang, Y-R Lin, S-C Liu, M D Burrow, S P Kowalski, C S Katsar, T A DelMonte, K A Feldmann, K F Schertz & J F Wendel

Non-canonical introns are at least 10⁹ years old 383

H-J Wu, P Gaubier-Comella, M Delseny, F Grellet, M Van Montagu & P Rouzé

Val92Met variant of the melanocyte stimulating hormone receptor gene 384

X Xu, M Thornwall, L-G Lundin & V Chhajlani

progress Genes responsible for human hereditary deafness: symphony of a thousand 385

Christine Petit

articles Cloning and characterization of a novel *bicoid*-related homeobox transcription factor gene, *RIEG*, involved in Rieger syndrome 392

E V Semina, R Reiter, N J Leysens, W L M Alward, K W Small, N A Datson, J Siegel-Bartelt, D Bierke-Nelson, P Bitoun, B U Zabel, J C Carey & J C Murray

Susceptible chiasmate configurations of chromosome 21 predispose to non-disjunction in both maternal meiosis I and meiosis II 400

N E Lamb, S B Freeman, A Savage-Austin, D Pettay, Lisa Taft, J Hersey, Y Gu, J Shen, D Saker, K M May, D Avramopoulos, M B Petersen, A Hallberg, M Mikkelsen, T J Hassold & S L Sherman **ncv**

Spontaneous X chromosome MI and MII nondisjunction events in *Drosophila melanogaster* oocytes have different recombinational histories 406

K E Koehler, C L Boulton, H E Collins, R L French, K C Herman, S M Laceyfield, L D Madden, C D Schuetz & R S Hawley **ncv**

Suppression of the novel growth inhibitor p33^{ING} promotes neoplastic transformation 415

I Garkavtsev, A Kazarov, A Gudkov & K Riabowol

Nature Genetics (ISSN 1061-4036) is published monthly by Nature Publishing Co., headquartered at 345 Park Avenue South, New York, NY 10010, which is owned by Nature America Inc., a subsidiary of Macmillan Magazines Ltd, of London. Editorial Office: Nature Genetics, 545 National Press Building, Washington DC 20045. Telephone (202) 626 2513. Fax (202) 626 0970. e-mail: natgen@naturedc.com. North American Advertising: Nature Genetics, 345 Park Avenue South, New York, NY 10010. Telephone (212) 726 9200. Fax (212) 696 9606. European Advertising: Nature Genetics, Porters South, Cannon Street, London, N1 9XW, UK. Telephone (0171) 833 4100. Fax (0171) 833 4101. e-mail: natgen@porters.com. New subscriptions, renewals, changes of address, back issues and all customer service questions should be addressed to: Nature Genetics Subscription Dept., 345 Park Avenue South, New York, NY 10010. Telephone (212) 726 9200. Fax (212) 696 9606. Outside North America: Nature Genetics, Macmillan Magazines Ltd, Porters South, Cannon Street, London N1 9XW, UK. Telephone (0171) 833 4000. Fax (0171) 843 4596. Annual subscription rates: US/Canada: US\$495 (institutional/corporate), US\$195 (individual making personal payment), Canada add 7% for GST. BN 14091 1595 RT. UK: £350 (institutional/corporate), £175 (individual making personal payment). Japan: Contact Nature Japan, K.K., Shin-Mitsuke Bldg, 4F, 3-6 Ichigaya Tamachi, Shinjuku-ku, Tokyo 162. Reprints: Nature Genetics Reprints Dept., 345 Park Avenue South, New York, NY 10010. Tel: (212) 726 9200. Fax: (212) 696 9606. Periodicals postage paid at New York, NY, and at additional mailing offices. Postmaster: send address changes to Nature Genetics Subscription Dept., 345 Park Avenue South, New York, NY 10010. Executive Officers of Nature America, Inc.: Nicholas Byam Shaw, Chairman of the Board; Mary Waltham, President; Edward Vais, Secretary-Treasurer. Published in Japan by Nature Japan KK, Shin-Mitsuke Building, 3F, 3-6 Ichigaya Tamachi, Shinjuku-ku, Tokyo 162, Japan. Printed in the USA by Cadmus Journal Services. ©1996 Nature America, Inc.

Nature Publishing Co.
345 Park Avenue South
10th floor
New York, NY 10010-1707
Tel: (212) 726-9200
Fax: (212) 696-9606

President-Publisher
Mary Waltham

Vice President Sales
Marion Delaney

Vice President Marketing
James A. Skowrenski

American Advertising Sales
Manager
Sande T. Giaccone (New York)

European Advertising Sales
Manager
Kathryn Wayman (London)

Classified Advertising Sales
Manager
Erika A. Simon (New York)
Mike Grant (London)

Assistant Classified Sales
Manager
Benjamin Crowe (New York)

Production & Information
Systems Director
Nick Kemp

Circulation Manager
Maira Musto (New York)
Nic Harman (London)

Group Marketing Manager
Lina Dzurenda



Macmillan Magazines Ltd
Porters South
Crinan Street
London N1 9XW
Tel: 44 (0)171 833 4000
Fax: 44 (0)171 843 4596

Managing Director
Ray Barker

Publishing Director
Andy Sutherland

Editor-in-Chief,
Nature publications
Philip Campbell

Art Director
Jane Walker

Nature Japan KK
Shin-Mitsuke Bldg
3-6 Ichigaya Tamachi
Shinjuku-ku
Tokyo 162
Telephone 03 3267 8751
Fax 03 3267 8746

articles

- A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants** 421
D Leister, A Ballvora, F Salamini & C Gebhardt
- Identification of a RING protein that can interact *in vivo* with the BRCA1 gene product** 430
L C Wu, Z W Wang, J T Tsan, M A Spillman, A Phung, X L Xu, M-C W Yang, L-Y Hwang, A M Bowcock & R Baer
- Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis** 441
J G Hacia, L C Brody, M S Chee, S P A Fodor & F S Collins
- Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy** 450
D D Shoemaker, D A Lashkari, D Morris, M Mittmann & R W Davis

letters

- Use of a cDNA microarray to analyse gene expression patterns in human cancer** 457
J DeRisi, L Penland & P O Brown (Group 1); M L Bittner, P S Meltzer, M Ray, Y Chen, Y A Su & J M Trent (Group 2)
- Retinal-specific guanylate cyclase gene mutations in Leber's congenital amaurosis** 461
I Perrault, J M Rozet, P Calvas, S Gerber, A Camuzat, H Dollfus, S Châtelain, E Souied, I Ghazi, C Leowski, M Bonnemaïson, D Le Paslier, J Frézal, J-L Dufier, S Pittler, A Munnich & J Kaplan
- Complex interactions of new quantitative trait loci, Sluc1, Sluc2, Sluc3, and Sluc4, that influence the susceptibility to lung cancer in the mouse** 465
R J A Fijneman, S S de Vries, R C Jansen & P Demant
- Gene interaction and single gene effects in colon tumour susceptibility in mice** 468
T van Wezel, A P M Stassen, C J A Moen, A A M Hart, M A van der Valk & P Demant
- A major quantitative trait locus influences hyperactivity in the WKHA rat** 471
M-P Moisan, H Courvoisier, M-T Bihoreau, D Gauguier, E D Hendley, M Lathrop, M R James & P Mormède
- An H-YD^b epitope is encoded by a novel mouse Y chromosome gene** 474
A Greenfield, D Scott, D Pennisi, I Ehrmann, P Ellis, L Cooper, E Simpson & P Koopman
- Homozygosity mapping of Hallervorden-Spatz syndrome to chromosome 20p12.3-p13** 479
T D Taylor, M Litt, P Kramer, M Pandolfo, L Angelini, N Nardocci, S Davis, M Pineda, H Hattori, P J Flett, M R Cilio, E Bertini & S J Hayflick
- Identification of BTG2, an antiproliferative p53-dependent component of the DNA damage cellular response pathway** 482
J-P Rouault, N Falette, F Guéhenneux, C Guillot, R Rimokh, Q Wang, C Berthet, C Moyret-Lalle, P Savatier, B Pain, P Shaw, R Berger, J Samarut, J-P Magaud, M Ozturk, C Samarut & A Puisieux

correction/errata

See pages 487-488

Use of a cDNA microarray to analyse gene expression patterns in human cancer

Joseph DeRisi¹*, Lolita Penland² & Patrick O. Brown² (Group 1); Michael L. Bittner³*, Paul S. Meltzer³, Michael Ray³, Yidong Chen³, Yan A. Su³ & Jeffrey M. Trent³ (Group 2)

The development and progression of cancer¹⁻³ and the experimental reversal of tumorigenicity^{4,5} are accompanied by complex changes in patterns of gene expression. Microarrays of cDNA provide a powerful tool for studying these complex phenomena⁶⁻⁸. The tumorigenic properties of a human melanoma cell line, UACC-903, can be suppressed by introduction of a normal human chromosome 6, resulting in a reduction of growth rate, restoration of contact inhibition, and suppression of both soft agar clonogenicity and tumorigenicity in nude mice^{4,5,9}. We used a high density microarray of 1,161 DNA elements to search for differences in gene expression associated with tumour suppression in this system. Fluorescent probes for hybridization were derived from two sources of cellular mRNA [UACC-903 and UACC-903(+6)] which were labelled with different fluors to provide a direct and internally controlled comparison of the mRNA levels corresponding to each arrayed gene. The fluorescence signals representing hybridization to each arrayed gene were analysed to determine the relative abundance in the two samples of mRNAs corresponding to each gene. Previously unrecognized alterations in the expression of specific genes provide leads for further investigation of the genetic basis of the tumorigenic phenotype of these cells.

DNA microarrays, containing 1,161 total elements, including 870 different cDNAs and controls⁹⁻¹¹ (see Methods), were printed robotically onto a glass microscope slide in four quadrants covering an area of about 1 cm² (Fig. 1). We prepared fluorescent cDNA probes using total poly (A)⁺ mRNA from UACC-903 cells and UACC-903(+6) cells by labelling with a green and red fluor, respectively. A mixture of the two fluorescently labelled probes was hybridized to the DNA microarray. This comparative hybridization method, coupled with the doping of synthetic standards and an estimation of statistically significant deviation for local background variance allowed a direct and quantitative comparison of the relative abundance of individual DNA sequences in this complex sample⁶⁻⁸. We added a set of synthetic poly (A)⁺-tailed 'mRNAs' to the purified mRNA from each cell line as internal standards to assist in quantita-

tion. In addition to the synthetic standards, we added complementary to genes preferentially expressed in the tumorigenic UACC-903 cell line, and the reddish spots correspond to genes preferentially expressed in the non-tumorigenic UACC-903(+6) cell line. Genes expressed at approximately equal levels in the two cell lines appear yellow or brown. A portion of the array at higher magnification highlights the diverse pattern of differential expression observed (Fig. 2b). In Fig. 2a, rectangles corresponding to specific array elements are coloured to reproduce the hue and intensity of the fluorescent signal at each element. The hybridization signals from a duplicated set of genes are shown juxtaposed, to illustrate the reproducibility of the hybridization signals for each gene.

To address the possibility that an apparent difference in expression might result from experimental variables unrelated to the difference in chromosomal composition between the two cell lines, we examined the variance in expression for 90 'housekeeping' genes. We selected these genes based on the assumption that they would not be differentially expressed between the two cell lines. The averaged red/green ratio for this subset of genes was 1.13. The averaged red/green ratio for the set of five internal standards was 0.97 ($n = 10$). The variability in the expression level of the housekeeping genes probably overestimates the experimental variability in measuring differential expression. As a conservative standard, an absolute fluorescent signal (red or green) with an intensity greater than that observed at the control array elements containing total human genomic DNA was considered to represent specific hybridization. Gene-specific hybridization was therefore only considered significantly different between samples if the following two criteria were met: i) the signal intensity (green or red) exceeded this threshold; and ii) the logarithm of the red/green fluorescence signal ratio differed by ≥ 3 S.D. from the mean logarithm of this ratio for the 'housekeeping' gene panel (that is, ratios < 0.52 or > 2.4).

By these criteria, mRNA levels for 15/870 (1.7%) genes were significantly diminished, while the mRNA levels for 63/870 (7.3%) genes were significantly increased in association with suppression of tumorigenicity by introduction of chromosome 6. To test the reliability of microarray hybridization results in identifying differentially expressed genes, we analysed 16 genes by northern analysis. In each case, the results of northern analysis corroborated the differential gene expression identified by microarray hybridization (Fig. 3).

Significant differences in expression between these two cell lines identified several genes as candidates for determining features of the tumorigenic phenotype of the melanoma cells. For example, among the genes detected with significantly higher expression (> 10 -fold) in the tumorigenic cells was the human brown locus protein (TRP1/melanoma antigen gp75). This is the most abundant glycoprotein in melanocytic cells and a critical melanosome membrane protein^{12,13}. Additionally, its

¹Howard Hughes Medical Institute, ²Department of Biochemistry, Stanford University Medical Center, Stanford, California 94305, USA

³Laboratory of Cancer Genetics, National Center for Human Genome Research, National Institutes of Health, Bethesda, Maryland 20892, USA

*J.D. & M.L.B. contributed equally to this work.

Correspondence should be addressed to P.O.B. or J.T. e-mail: pbrown@cngm.stanford.edu; trent@chgr.

Complementary to these standards were included, to duplicate, on the microarray. Based on these standards, mRNA species comprising 1:10,000 of the mass of the poly (A)⁺ RNA could readily be detected.

In a representative two-colour fluorescent scan of all

array elements, the expression of the TRP1 gene was significantly higher in the tumorigenic cells than in the non-tumorigenic cells. This result was confirmed by northern analysis. In addition, the expression of the TRP1 gene was significantly higher in the tumorigenic cells than in the non-tumorigenic cells. This result was confirmed by northern analysis. In addition, the expression of the TRP1 gene was significantly higher in the tumorigenic cells than in the non-tumorigenic cells. This result was confirmed by northern analysis.

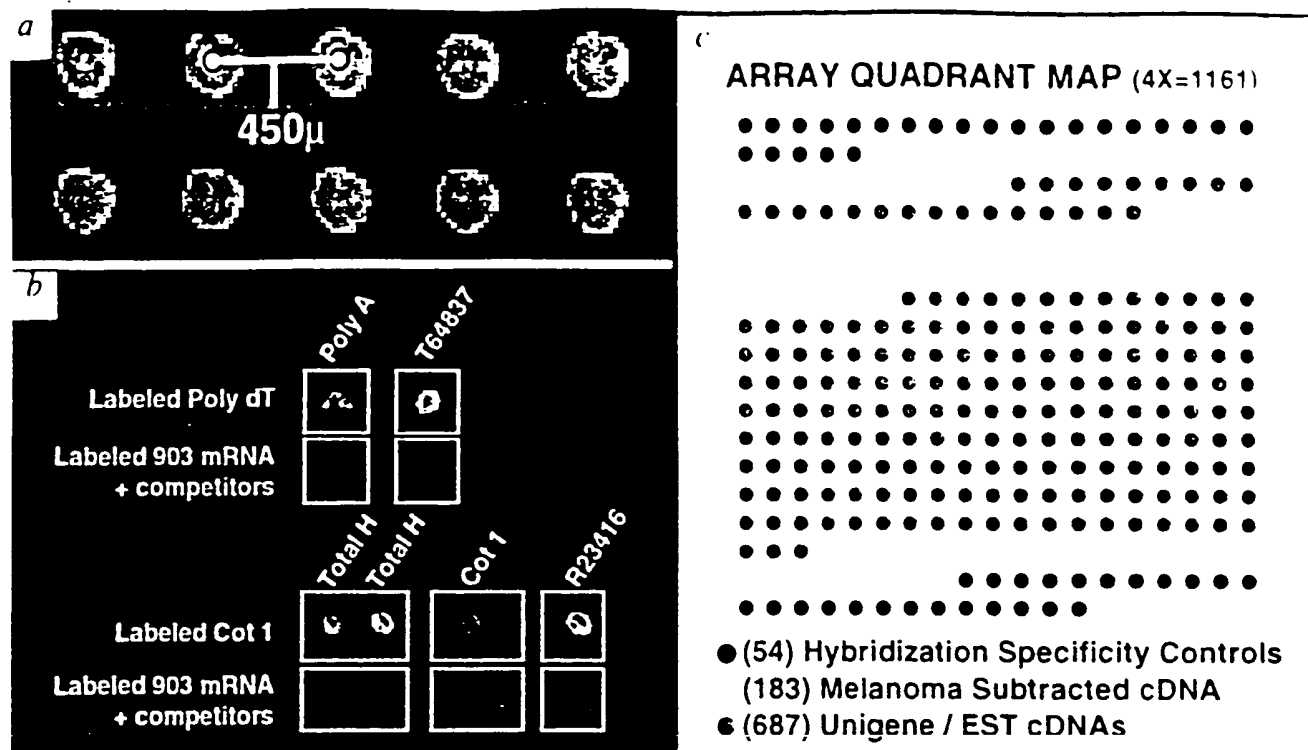


Fig. 1 Properties of cDNA microarrays. **a**, A fluorescent scan of DNA printed onto a poly-lysine coated slide. The DNA is stained with a DNA-specific fluorescent dye, YOYO. The center-to-center spacing of adjacent spots is 450 μ , allowing the potential for up to 10,000 spots/2.54 X 7.62 cm microscope slide. **b**, Efficient blocking of hybridization to DNA repeats. Hybridization of fluorescein-labelled poly (dT)* to arrays in the absence of competitor produces strong hybridization to immobilized poly (dA)* as well as to some cDNAs, such as the EST T64827 shown. Rhodamine-labelled cDNA (red) from the UACC-903 cell line hybridized in the presence of poly (dA)* blocker shows little if any signal at either site (Total H = total human). Similarly, hybridization with fluorescein-labelled Cot1 DNA in the absence of competitor produces bright signal on immobilized Cot1 DNA, total human DNA and at some cDNA elements (presumed to contain highly repeated sequences, such as R23416); while Rhodamine-labelled cDNA (red) from the UACC-903 cell line produces little if any signal at these locations when hybridized in the presence of excess unlabelled poly (dA)*, and human Cot1 DNA. The absence of signal at some cDNA locations following UACC-903 cDNA hybridizations also indicates that the PCR-amplified, plasmid vector sequences at all cDNA targets do not contribute significant hybridization signal. **c**, Schematic of the array organisation. Robotic printing from 96 well microtiter trays was carried out with 4 print heads, spaced to fit into 4 adjacent microtiter wells. This maps the contents of each tray into four separate quadrants on the glass slide. A colour-coded map of the general distribution of target types in each of the resulting quadrants is shown.

els were elevated by the addition of a normal chromosome 6 (17 genes) are known to be activated by IFN- γ , a cardinal proinflammatory cytokine that, among other activities, induces expression of the gene products of the MHC class II locus. For example, the mRNA encoding monocyte chemotactic protein 1 (MCAF/MCP1), a cytokine that induces monocyte chemotaxis and activation^{15,16}, was more than 10-fold less abundant in the tumorigenic cell line. In the skin, MPC1 is critical in the regulation of cutaneous monocyte trafficking¹⁶⁻¹⁷, and elevated expression plays a role in suppression of tumour growth and metastasis¹⁹⁻²¹. The mechanism by which these interferon- γ regulated genes are induced in UACC-903 cells by transfer of a normal chromosome 6 remains to be determined. It is worth noting, however, that the interferon- γ receptor gene is localized to the distal long arm of human chromosome 6.

Finally, several genes that showed >10-fold higher expression in the suppressed UACC-903(+6) cells have previously been recognized in other models of tumour suppression. Most notably, there was elevated expression of the mRNA encoding WAF1 (p21), a key mediator of tumour suppression by p53 (ref. 18). The p21

These results provide a wide view of the diverse systems that are altered in this model system of tumorigenicity, and focus attention on specific gene products and pathways that may be of particular importance in this tumour type.

Our ability to classify human cancers in a way that reflects the underlying molecular pathology or that anticipates their potential for progression or response to treatment, remains primitive. Using cDNA microarrays to define alterations in gene expression associated with a specific cancer may be an efficient way to uncover clues to the specific molecular derangements that contribute to its pathogenesis and thus identify potential targets for therapeutic intervention. Moreover, recognition of pathognomonic alterations in gene expression might provide a basis for improved diagnosis and molecular classification of cancers and thus allow selection of the most appropriate therapeutic strategies.

Public databases of human expressed gene sequences contain partial sequences of at least 40,000 different human genes¹¹, and efforts to develop a human transcript map have developed rapidly²¹. Based on the high yield of information obtained using an array of <1,000 genes, we have shown that a single, complete array of human genes, which likely provide a rich source of new and useful insights into human biology and a deeper understanding of the gene pathways involved in the pathogenesis of cancer and other diseases.

transcriptional activation pathway. In normal melanocytes, melanoma cells are suppressed for metastasis by the introduction of chromosome 6, expression of WAF1 (p21) mRNA and protein correlates inversely with metastatic potential²⁰.

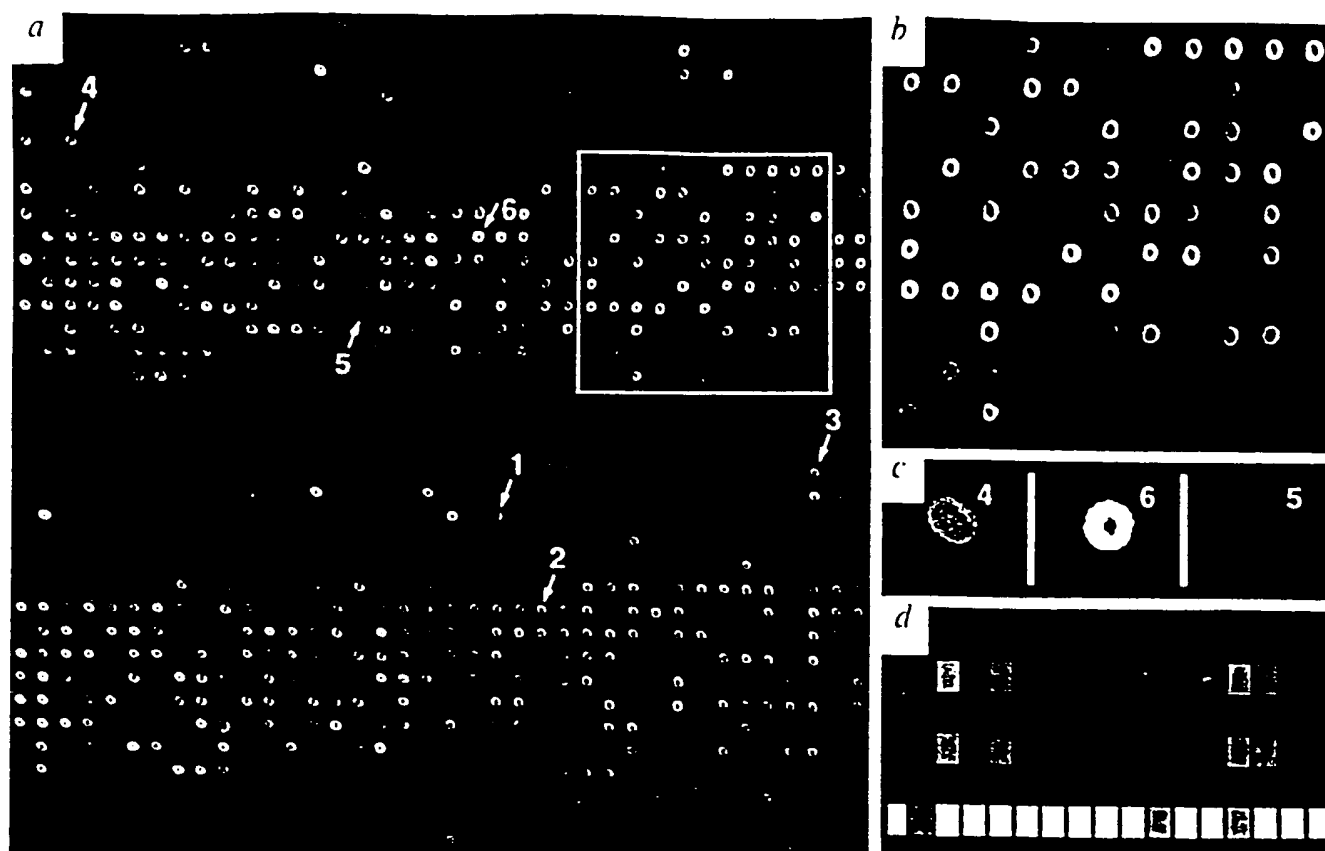


Fig. 2 DNA microarray analysis of changes in gene expression between the tumorigenic cell line, UACC-903, and its non-tumorigenic derivative, UACC903(+6), derived by introduction of a normal chromosome 6. **a**, A ratio image of the results of simultaneous hybridization of Rhodamine110-labelled cDNA (green) from UACC-903 and Cy3-labelled cDNA (orange-red) from UACC-903(+6) to a microarray. To produce this image, the scan images corresponding to each fluorescent probe were combined as the appropriate colour channels in a single image. Arrows indicate the location within the array of the corresponding genes analysed by northern blotting (Fig. 3). **b**, A magnified image of the area of the array boxed in white in (a). **c**, Magnified image of three cDNAs identified by arrows in (a), representing the cDNAs for: left, *MCAF/MCP-1* (*r/g* ratio >10); centre, *β -actin* (*r/g* ratio 1.04); and right, *α -1-antichymotrypsin* (*r/g* ratio 0.2) [see Fig. 3]. **d**, simplified representation of ratio hybridization results. Quantitative fluorescence intensity data is extracted from each array target. The average target colour ratio determines the hue of each box and the average intensity determines the brightness of each box. In this image, the order of the boxes corresponds to their original order in the microtiter plate from which they were printed. Duplicate printings of the same plate can be examined side by side, as in the first two rows shown here, to assess reproducibility of the hybridization results (see text). Numbered arrows indicate the location within the array corresponding to genes analysed by northern blotting in Fig. 3.

Methods

Generation of microarrays, hybridization, scanning. The preparation of coated microscope slides and subsequent robotic printing of DNA was carried out in a manner similar to that described¹. Briefly, pre-cleaned glass slides were treated with poly-L-lysine solution (Sigma) to form an adhesive surface for printing. PCR products, purified by ethanol purification, were resuspended in 3 \times SSC. A custom built arraying robot picked up and deposited small volumes (\sim 5 nanoliters) of DNA onto the slides. After printing, the slides were washed in a 0.2% SDS solution. The remaining bound DNA was denatured by submerging the slides in 95 °C distilled water for 2 min followed by a brief wash with 95% ethanol. DNA was UV crosslinked to the slides (Stratagene Stratalinker, 60 mJ). To prevent non-specific probe binding, the slides were blocked by rinsing in a solution of 70 mM succinic anhydride dissolved in 0.1 M boric acid pH 8.0, containing 35% 1-methyl-2-pyrrolidone (Aldrich). Additional protocols and parts list pertaining to microarray fabrication can be obtained from <http://cmgm.stanford.edu/obrowe>.

by N. Ziv. A separate scan, using the appropriate excitation line, was done for each of the two fluorophores used. Data was collected at a maximum resolution of 9 microns/pixel with 12 bits of depth.

Probe preparation and labelling. RNA was extracted from cells using the Triazol reagent (LTI Inc.), following the manufacturer's directions. cDNA probes were synthesized from singly oligo dT-selected (Pharmacia) mRNA pools. Fluorescently labelled cDNA was prepared from mRNA by oligo dT-primed polymerization using SuperScript II reverse transcriptase (LTI Inc.). The pool of nucleotides in the labelling reaction was 0.5 mM dGTP, dATP and dCTP and 0.2 mM dTTP. Fluorescent nucleotides, Rhodamine 110 dUTP (Perkin Elmer Cetus) or Cy3 dUTP (Amersham), were present at 0.1 mM. Probes were purified by gel chromatography (BioSpin 6/BioRad) and ethanol precipitation.

Selection of cDNA elements and generation of control templates. Synthetic cDNAs were prepared by cloning random

¹Obrowe, M. & Ziv, N. (1998) *Journal of Microarray Technology*, 9, 100-105. This work was supported by National Institutes of Health Grant R01 CA70000 to N. Ziv. We thank Dr. J. Schlesselman for providing the UACC-903 cell line and Dr. J. Schlesselman for providing the UACC-903(+6) cell line.

²Obrowe, M. & Ziv, N. (1998) *Journal of Microarray Technology*, 9, 100-105. This work was supported by National Institutes of Health Grant R01 CA70000 to N. Ziv. We thank Dr. J. Schlesselman for providing the UACC-903 cell line and Dr. J. Schlesselman for providing the UACC-903(+6) cell line.

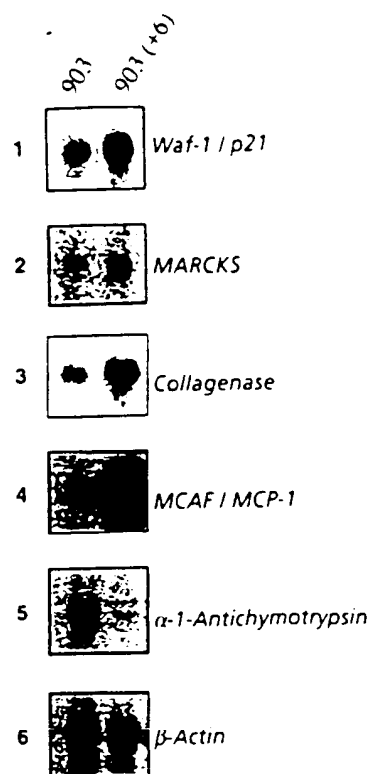


Fig. 3 Northern hybridization substantiating the consistency of the cDNA microarray results. Corresponding locations within the cDNA microarray illustrated in Fig. 2a are provided for 1) *Waf-1/p21*; 2) *MARCKS*; 3) *collagenase*; 4) *MCAF/MCP-1*; 5) α -1-antichymotrypsin; and 6) β -actin. The signal detected by a radio-labelled β -actin probe represents a control for loading variance, with a red/green ratio observed on the cDNA microarray (Fig. 2a,c) for β -actin of 1.04.

to the UniGene EST clustering system^{21,22}. The second largest group of clones consisted of 183 sequenced cDNA clones generated by subtraction of cDNA from the chromosome-6 suppressed non-tumorigenic UACC-903 (+6) cell line with cDNA from its parental tumorigenic cell line UACC-903 (ref. 9). Approximately 100 additional genes (total 870 genes arrayed) were obtained from EST libraries on the basis of their expression pattern (tissue specific, and so on). Each array included the following hybridization controls: plasmid vector, lambda, ϕ X174 phage, total human DNA, human *Colt* DNA, and poly (A)⁺. The synthetic standards used for normalization of signals in each wavelength were also arrayed. Controls were included in

each quadrant of the array to assess the reproducibility of the hybridization signal. Two plates of cDNA clones (derived from the UACC-903 subtracted library) were also arrayed in duplicate. Fidelity of the Unigene array relative to dbEST was tested by sequencing of a random sample of 11 clones used for microarray construction. All sequences were identical with the

corresponding dbEST entries. Additionally, each microarrayed cDNA from the UACC-903 subtracted library was sequenced. A listing of cDNAs comprising this microarray which were derived from the Unigene and 'housekeeping' panel can be obtained from <http://www.nih.gov/DIR/LCG/ARRAY/exprn.html>.

Northern blot analysis. Total RNA, 10 μ g per lane, was electrophoresed in 1.2% agarose-formaldehyde gels and transferred onto nylon membrane (Hybond-N⁺, Amersham) by capillary blotting overnight. For DNA probes insert fragments from the Soares INIB cDNA library¹⁰ were obtained by vector PCR for p21, MARCKS, α -1-antichymotrypsin and β -actin. Probes for fibroblast collagenase and MCAF/MCP-1 were isolated from a UACC-903(+6) enriched cDNA library⁹ with all probes labelled by random priming. Filters were washed to a stringency of 0.1x SSC at 42 °C for 20 min.

Web sites. <http://cmgm.stanford.edu/pbrown> for protocols and parts list pertaining to microarray fabrication. <http://www.ncbgr.nih.gov/DIR/LCG/ARRAY/exprn.html> for a listing of cDNAs comprising this microarray which were derived from the Unigene and 'housekeeping' panel.

Acknowledgements

Work in P.O.B.'s laboratory is supported in part by the Howard Hughes Medical Institute and National Center for Human Genome Research (HG00450). We would like to acknowledge the excellent technical and graphic assistance of X. He, T. Hofmann, Y. Jiang, J. Leuders, D. Leja and B. Walker. J.D. was supported by NIH grant 2T32BM07276-21. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

Received 15 October; accepted 8 November, 1996.

- Vogelstein, B. & Kinzler, K.W. The multistep nature of cancer. *Trends Genet.* 9, 138-141 (1993).
- Weinberg, R.A. The molecular basis of oncogenes and tumor suppressor genes. *Ann NY Acad. Sci.* 758, 331-338 (1995).
- Levine, A.J. The tumor suppressor genes. *Annu. Rev. Biochem.* 62, 623-651 (1993).
- Trent, J.M. et al. Tumorigenicity in human melanoma cell lines controlled by introduction of human chromosome 6. *Science* 247, 568-571 (1990).
- Su, Y. et al. Reversion of monochromosome-mediated suppression of tumorigenicity in malignant melanoma by retroviral transduction. *Cancer Res.* 56, 3186-3191 (1996).
- Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470 (1995).
- Shalon, D., Smith, S.J. & Brown, P.O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* 6, 639-645 (1996).
- Schena, M. et al. Parallel human genome analysis: microarray-based expression of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93, 10539-10544 (1996).
- Ray, M.E., Su, Y.A., Meltzer, P.S. & Trent, J.M. Isolation and characterization of genes associated with chromosome-6 mediated tumor suppression in human malignant melanoma. *Oncogene* 12, 2527-2533 (1996).
- Soares, M.B. et al. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* 91, 9228-9232 (1994).
- Boguski, M.S. & Schuler, G.D. ESTablishing a human transcript map. *Nature Genet.* 10, 369-371 (1995).
- Vijayasaradhi, S., Doskotch, P.M., Wolchok, J. & Houghton, A.N. Melanocyte differentiation marker gp75, the brown locus protein, can be regulated independently of tyrosinase and pigmentation. *J. Invest. Dermatol.* 105, 113-119 (1995).
- Vijayasaradhi, S., Xu, Y., Bouchard, B. & Houghton, A.N. Intracellular sorting and targeting of melanosomal membrane proteins: identification of signals for sorting of the human brown locus protein, gp75. *J. Invest. Dermatol.* 130, 807-820 (1995).
- Nakao, J. et al. Expression of proteolipid protein gene is directly associated with secretion of a factor influencing oligodendrocyte development. *J. Neurochem.* 64, 2396-2403 (1995).
- Graves, D.T., Barnhill, R., Galanopoulos, T. & Antoniadou, H.N. Expression of monocyte chemoattractant protein-1 in human melanoma in vivo. *Am. J. Pathol.* 140, 9-14 (1992).
- Kristensen, M.S., Deleuran, B.W., Larsen, C.G., Thstrup-Pedersen, K. & Paludan, K. Expression of monocyte chemoattractant and activating factor (MCAF) in skin related cells. A comparative study. *Cytokine* 5, 520-524 (1993).
- Huang, S., Xie, K., Singh, R.K., Gulman, M. & Bar-Eli, M. Suppression of tumor growth and metastasis of murine renal adenocarcinoma by syngeneic fibroblasts genetically engineered to secrete the JE/MCP-1 cytokine. *J. Interferon Cytokine Res.* 15, 655-665 (1995).
- El-Deiry, W.S. et al. WAF1, a potential mediator of p53 tumor suppression. *Cell* 75, 817-825 (1993).
- Miele, M.E. et al. Metastasis suppressed, but tumorigenicity and local invasiveness unaffected, in the human melanoma cell line MelJuSo after introduction of human chromosomes 1 or 6. *Mol. Carcinog.* 15, 284-299 (1996).
- Jiang, H. et al. The melanoma differentiation-associated gene mda-6, which encodes the cyclin-dependent kinase inhibitor p21, is differentially expressed during growth, differentiation and progression in human melanoma cells. *Oncogene* 10, 1855-1864 (1995).
- Schuler, G.D. et al. A gene map of the human genome. *Science* 274, 540-546 (1996).
- Lennon, G., Aufray, C., Polymeropoulos, M. & Soares, M.B. The IMAGE Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 33, 151-152 (1996).

GENOME RESEARCH

July 1996

Volume 6 Number 7

INCLUDING



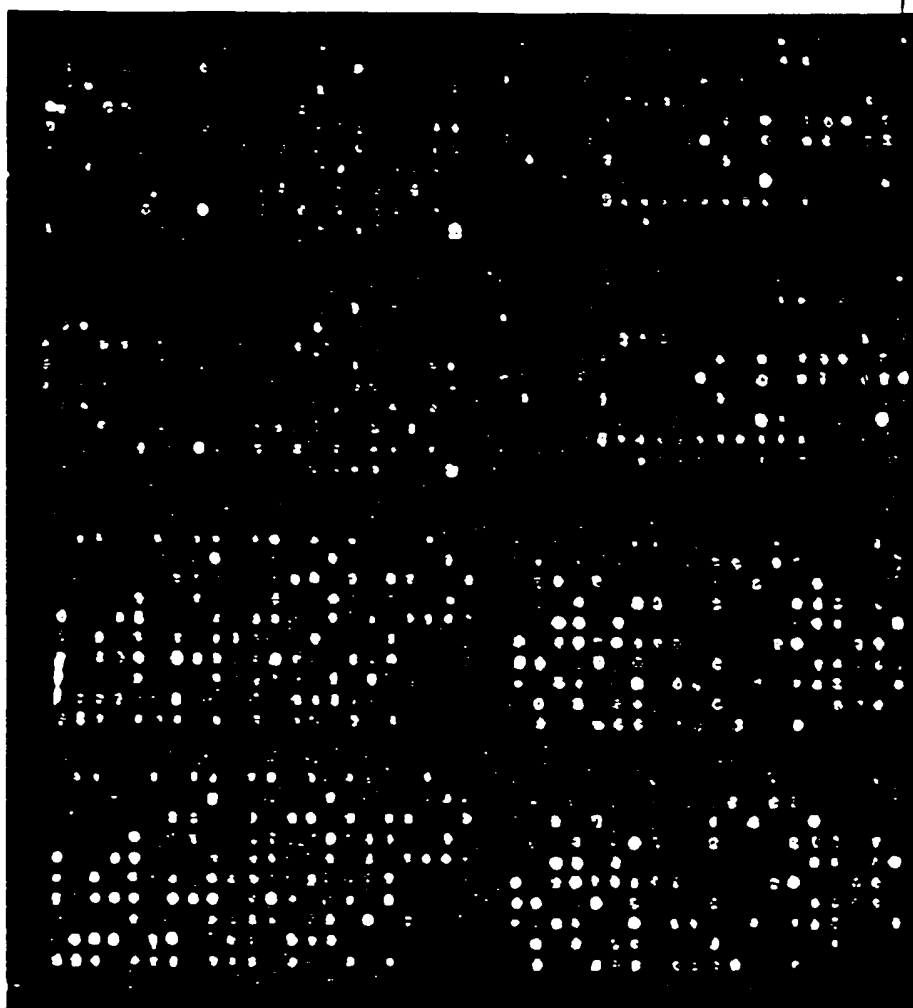
IBD Mapping in Livestock

Sequence of 500-kb
Rhizobium Replicon

Human Y Chromosome
Haplotypes

BAC Mapping of
Extrachromosomal Structure

DNA Microarray System





Advertise in
Genome Research
and reach the people doing
the most exciting science of the 90s!!

Please call or FAX Teresa Tiganis, Advertising Manager for further details.
Tel. (516) 367-8351, FAX (516) 367-8532.

Editorial office: Cold Spring Harbor Laboratory Press, 1 Bungtown Road, Cold Spring Harbor, New York 11724-2203. Phone 516-367-8492; FAX 516-367-8334.

GENOME RESEARCH (ISSN 1054-9803) is published monthly for \$495 (U.S. institutional; \$545 R.O.W.), \$95 (individual making personal payment; \$145 R.O.W., includes airlift) by Cold Spring Harbor Laboratory Press, 1 Bungtown Road, Cold Spring Harbor, New York 11724. Periodicals class postage pending is paid at

10 Skyline Drive, Plainview, New York 11803-2500. Telephone: Continental U.S. and Canada 1-800-843-4388; all other locations 516-349-1930. FAX 516-349-1946. Personal subscriptions must be prepaid by personal check, credit card, or money order. Claims for missing issues must be received within 4 months of issue date.

Manager. Personal: U.S. \$95; R.O.W. \$145 (includes airlift). Institutional: U.S. \$495; R.O.W. \$545 (includes airlift). Orders may be sent to Cold Spring Harbor Laboratory Press, Fulfillment Department, 10 Skyline Drive, Plainview, New York 11803-2500. Telephone: Continental U.S. and Canada 1-800-843-4388; all other locations 516-349-1930. FAX 516-349-1946. Personal subscriptions must be prepaid by personal check, credit card, or money order. Claims for missing issues must be received within 4 months of issue date.

Copyright information: Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Cold Spring Harbor Laboratory Press for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of \$5.00 per copy is paid directly to CCC, 21 Congress Street, Salem, Massachusetts 01970 (1054-9803/96 - \$5.00). This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional

Copyright information: Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Cold Spring Harbor Laboratory Press for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of \$5.00 per copy is paid directly to CCC, 21 Congress Street, Salem, Massachusetts 01970 (1054-9803/96 - \$5.00). This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional

Copyright © 1996 by Cold Spring Harbor Laboratory Press

RESEARCH PAPERS

- | | | |
|--|--|-----|
| Gene Transfer into Corn Earworm
(<i>Helicoverpa zea</i>) Embryos | James D. DeVault, Keith J. Hughes,
Roger A. Leopold, Odell A. Johnson,
and Sudhir K. Narang | 571 |
| Identity-by-descent Mapping of Recessive Traits
in Livestock: Application to Map the Bovine
<i>Syndactyly</i> Locus to Chromosome 15 | Carole Charlier, Frédéric Farnir,
Paulette Berzi,
Pascal Vanmanshoven,
Benoît Brouwers, Hans Vromans,
and Michel Georges | 580 |
| Sequencing the 500-kb GC-rich Symbiotic
Replicon of <i>Rhizobium</i> sp. NGR234 Using Dye
Terminators and a Thermostable "Sequenase":
A Beginning | Christoph Freiberg, Xavier Perret,
William J. Broughton, and
André Rosenthal | 590 |
| Worldwide Distribution of Human
Y-chromosome Haplotypes | Fabício R. Santos,
Néstor O. Bianchi, and
Sérgio D.J. Pena | 601 |
| Bacterial Artificial Chromosome Cloning and
Mapping of a 630-kb Human
Extrachromosomal Structure | Min Wang, Stephanie Shouse,
Barbara Lipes, Ung-Jin Kim,
Hiroaki Shizuya, and Eric Lai | 612 |

LETTERS

- | | | |
|--|--|-----|
| The Genomic Structure of Discoidin Receptor
Tyrosine Kinase | Martin P. Playford, Robin J. Butler,
Xiao Cun Wang, Roy M. Katso,
Inez E. Cooke, and
Trivadi S. Ganesan | 620 |
| A Contiguous High-resolution Radiation Hybrid
Map of 44 Loci from the Distal Portion of the
Long Arm of Human Chromosome 5 | Janet A. Warrington and
John J. Wasmuth | 628 |

GENOME METHODS

Uniform Amplification of a Mixture of Deoxyribonucleic Acids with Varying GC Content	Namadev Baskaran, Rajendra P. Kandpal, Ajay K. Bhargava, Michael W. Glynn, Allen Bale, and Sherman M. Weissmann	633
A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization	Dari Shalon, Stephen J. Smith, and Patrick O. Brown	639
Microsatellite Hybrid Capture Technique for Simultaneous Isolation of Various STR Markers	Michal Prochazka	646
Erratum		650

Product News	651
--------------	-----

COVER DNA microarrays for analyzing complex DNA samples. Shown is a two-color fluorescent scan of an 1.8-cm × 1.8-cm yeast array of λ clones of yeast genomic DNA. (For details, see Shalon et al., p. 639.)

A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization

Dari Shalon,^{1,4} Stephen J. Smith,³ and Patrick O. Brown^{1,2,5}

¹Howard Hughes Medical Institute and Departments of ²Biochemistry and ³Molecular and Cellular Physiology, Stanford University, Stanford, California 94305

Detecting and determining the relative abundance of diverse individual sequences in complex DNA samples is a recurring experimental challenge in analyzing genomes. We describe a general experimental approach to this problem, using microscopic arrays of DNA fragments on glass substrates for differential hybridization analysis of fluorescently labeled DNA samples. To test the system, 864 physically mapped λ clones of yeast genomic DNA, together representing >75% of the yeast genome, were arranged into 1.8-cm \times 1.8-cm arrays, each containing a total of 1744 elements. The microarrays were characterized by simultaneous hybridization of two different sets of isolated yeast chromosomes labeled with two different fluorophores. A laser fluorescent scanner was used to detect the hybridization signals from the two fluorophores. The results demonstrate the utility of DNA microarrays in the analysis of complex DNA samples. This system should find numerous applications in genome-wide genetic mapping, physical mapping, and gene expression studies.

Many problems in genome analysis depend on determining what specific sequences are represented in a complex DNA or RNA sample and at what abundance, for example, what genes are represented in a specific chromosome band or YAC clone, what intervals are amplified or deleted in a particular cancer cell, or what genes are expressed in specific cells under specific conditions. As a general approach to this problem, we have developed a system for making microarrays of DNA samples on glass substrates, probing them by hybridization with complex fluorescent-labeled probes, and using a laser-scanning microscope to detect the fluorescent signals representing hybridization. Fluorescent labeling allows for simultaneous hybridization and separate detection of the hybridization signal from two or more probes. This in turn allows very accurate and reliable measurement of the relative abundance of specific sequences in two complex samples.

RESULTS

Array Hybridization Pattern

Figure 1 shows the two-color fluorescent scan of a yeast genomic array following hybridization

with a mixed probe consisting of lissamine-labeled DNA from the 6 largest yeast chromosomes together with fluorescein-labeled DNA from the 10 smallest yeast chromosomes. A red color indicates that yeast sequences present in the lissamine-labeled hybridization probe hybridized to an array element. A yellow-green color indicates that yeast sequences present in the fluorescein-labeled hybridization probe hybridized to an array element. An orange color indicates cross-hybridization of both chromosome pools to an array element (e.g., dispersed repetitive elements, such as Ty1 elements).

Each clone was spotted twice, resulting in duplicate hybridization patterns in adjacent quadrants of the array. Control DNA spots, which were randomly amplified in the same manner as the λ clone array elements, are located in the bottom corner of each quadrant. "A" points to a pair of spots containing total yeast genomic DNA. These spots appear orange because both chromosome pools hybridized to yeast genomic DNA. The negative controls are as follows: "B" points to a pair of spots of wild-type λ DNA, "C" points to a pair of human genomic DNA spots, and "D" points to a pair of λ 174 DNA spots. The lack of hybridization signals in these three negative control spots indicates that the hybridization was specific for yeast sequences.

Present address: Syntex Inc., Palo Alto, California 94304.
Corresponding author.
E-MAIL: pbrown@cmgm.stanford.edu, <http://cmgm.stanford.edu/pbrown>, FAX (415) 723-1399.

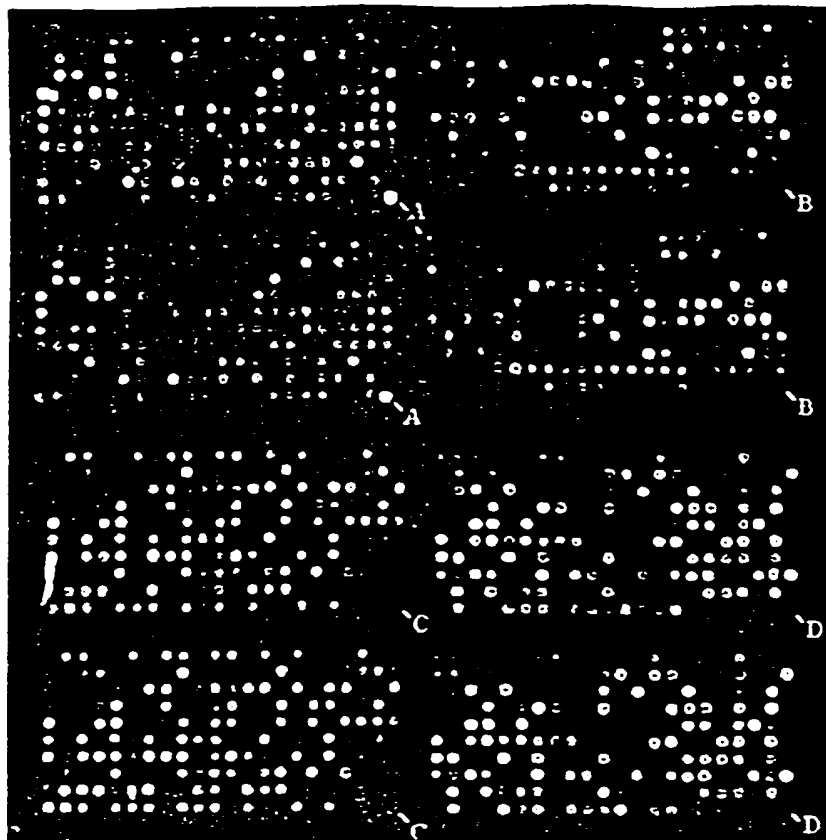


Figure 1 Two-color fluorescent scan of a 1.8-cm \times 1.8-cm yeast array of λ clones of yeast genomic DNA. The DNA spots are spaced at a distance of 380 μ m from center to center. A probe mixture consisting of DNA from the 6 largest yeast chromosomes (4, 7, 12, 13, 15, 16) labeled with lissamine (red dots) and DNA from the 10 smallest yeast chromosomes (1, 2, 3, 5, 6, 8, 9, 10, 11, 14) labeled with fluorescein (yellow-green dots) was hybridized to the array. A pair of yeast genomic DNA spots (A) served as a positive control. The three negative controls are λ DNA (B), human genomic DNA (C), and ϕ X174 DNA (D).

Karyotype Depiction of the Array Hybridization Pattern

The inserts contained in the arrayed λ clones have been mapped physically (Riles et al. 1993). The clones are arrayed in a random but known order on the array. Therefore, using the identity of each clone along with its physical map information, the pattern of hybridization to the yeast array can be represented in the form of a karyotype of the yeast genome, as shown in Figure 2. The color of any segment of the ideogram representing an individual chromosome on the karyotype is directly determined by the ratio of red and yellow-green fluorescent clones. The length of each discrete colored segment of each chromosome correspond to the physical lengths of the yeast

inserts. The chromosome segments colored black represent either intervals of the genome that are not represented by clones in the library (90%) or false-negative hybridization signals on the array (10%). Most of these false negatives are attributable to failures of the PCR amplification of the λ clones, though occasional failures of the arraying process or nonuniform surface preparation could account for a small fraction of the false-negative signals. The large gap on chromosome 12 is the region coding for ribosomal DNA that was not represented among the arrayed clones. Genomic intervals represented by overlapping clones were assigned a color based on the hybridization signals of only one of the overlapping clones, chosen at random.

Note that in this representation of a yeast karyotype, the largest six chromosomes are mainly colored red. This indicates that most of the arrayed clones that were mapped previously to these six large chromosomes hybridized primarily to the lissamine-labeled probe prepared from the corresponding purified chromosomes. Conversely, the smallest 10 chromosomes are mainly colored green in this image, matching the original CHEF gel isolation of the chromosomes used as the hybridization probe. The

experiment was repeated with the yeast genome split into six discrete chromosome pools containing 2–4 chromosomes per pool using CHEF gel electrophoresis. The chromosomes in each pool were extracted from the gel, amplified, and fluorescently labeled. The six chromosome pools were hybridized to six separate yeast arrays. Forty-four λ clones gave a positive hybridization signal on all six arrays indicating that they contain yeast repetitive sequences (data not shown). These 44 clones and 10 clones with very weak hybridization signals were not included in the karyotype representation shown in Figure 2.

Some of the 100 arrayed clones, which appear in this karyotype representation as green bands on the otherwise red chromosomes or red

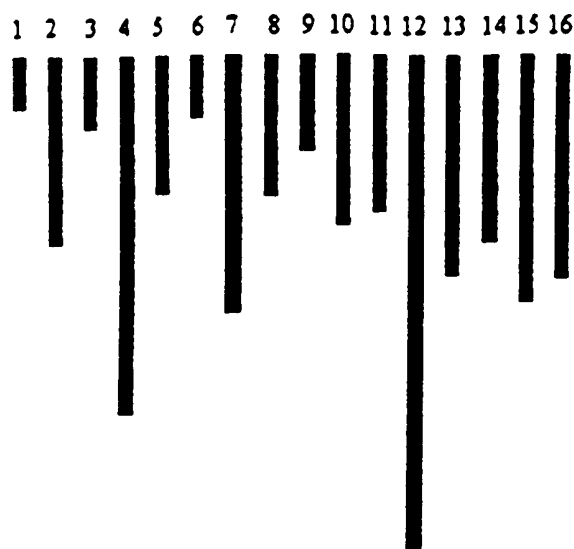


Figure 2 Computer-generated ideogram representing a karyotype of *S. cerevisiae*, based on the normalized hybridization signals from the array shown in Fig. 1. Note that the 6 largest chromosomes are mainly red and the 10 smallest chromosomes are mainly green. Black stripes represent intervals not represented by clones in the array or for which the corresponding clones gave false-negative hybridization signals.

bands on the otherwise green chromosomes. Four randomly chosen examples of these anomalous clones were analyzed by hybridizing the clones to vertical strips cut from a Southern blot of CHEF gel-separated yeast chromosomes. In each case, the hybridization patterns of the anomalous clones corroborated the chromosomal locations assigned by the microarray hybridization results (data not shown). Two clones that were thought to map to the 10 smallest chromosomes were found to hybridize preferentially to the probe representing the 6 largest chromosomes and thus appear as anomalous red bands on the karyotype. Both hybridized to one of the six largest chromosomes on the Southern blot. Similarly, two clones that appear as anomalous green bands on the karyotype were found to hybridize to one of the 10 smallest chromosomes on the Southern blot. Thus, the anomalous clones are probably the result of sample tracking errors or, possibly, of errors in the published restriction-digest-based physical map on which the karyotype representation was based (Riles et al. 1993).

similar to fluorescent in situ hybridization (FISH) to metaphase chromosomes, with three important differences. First, the target elements of the microarrays can, in principle, be any length or composition, from megabase YAC clones or microdissected chromosome bands to individual cDNA clones, to short oligonucleotides. This versatility allows the user to choose characteristics, such as the mapping resolution and genetic complexity of each array element, to suit a particular application. Second, the hybridization signals are localized to discrete elements of known size and location, making them easier to identify and quantitate than the hybridization signals from irregularly shaped metaphase spreads. Third, microarrays are more consistent and potentially amenable to automated production, hybridization, and data analysis than metaphase spreads.

Arrays of DNA samples on porous membranes, for example, dot blots, have long been used as a basic tool in molecular biology. Dot-blot membranes are usually at least 8×12 cm in size, require the use of milliliter volumes of hybridization solution, and are limited, owing to autofluorescence and scattering, to radioactive, chemiluminescent, and colorimetric hybridization detection methods (Ross et al. 1992). Microarrays made on glass surfaces, on the other hand, can be mass-produced and are comparatively inexpensive, convenient, and compatible with fluorescent hybridization detection methods. Furthermore, a glass surface, when appropriately treated, has very low nonspecific binding of labeled hybridization probes, resulting in lower backgrounds than are encountered typically with porous membranes. For hybridizations with very complex probes, the concentration of the labeled probe DNA is a limiting factor in the sensitivity of the assay. Minimizing the volume of the probe solution in a hybridization, by restricting the target to a small area and by using a nonporous substrate, makes it practical to achieve very high probe concentrations.

One important advantage of fluorescently labeled probes is that, unlike most radioactive and chemiluminescent signals, fluorescent signals do not disperse and therefore allow for very dense array spacing. A unique, and probably the most important, advantage of fluorescent probes is that the hybridization signals from two or more different probes can be detected simultaneously. In this way, two-color hybridization detection allows for a direct and quantitative comparison of the

The DNA microarray hybridization system reported here is conceptually and functionally

abundance of specific sequences between two probe mixtures that are hybridized competitively to a single array. The absolute intensity of a hybridization signal at a particular element in an array can vary owing to experimental factors such as variations in the amount of DNA deposited on the array, variations in the hybridization or wash conditions between experiments, or variations in the hybridization characteristics of the different DNA sequences on the array. The ratio of the two signals at any element in an array, however, is relatively insensitive to these confounding factors because they affect both probe mixtures equivalently. This ratio therefore accurately reflects the relative abundance of the cognate sequence in the two probe samples. This is the principle underlying the technique of comparative genomic hybridization (CGH), which is used to detect changes in the copy number of specific chromosomes or chromosomal regions (Kallioniemi et al. 1992). CGH is based on measuring the relative fluorescent hybridization intensities of two genomic-complexity hybridization probes, for example, probes representing genomic DNA from normal and affected tissue samples, which are labeled with two distinct fluorophores and hybridized simultaneously to a metaphase spread. DNA microarray representations of the human genome may provide a more convenient and higher resolution alternative to metaphase chromosomes for CGH.

Cross-hybridization between related sequences is an important problem faced by any hybridization-based assay, including the DNA microarray assay described here. Studies are now in progress to quantitate the extent of cross-hybridization between related sequences of varying homology and length, in DNA microarray hybridizations. The stringency of hybridization and washing can be controlled by varying the salt concentration and temperature as in conventional membrane-based hybridizations. Cross-hybridization caused by repetitive sequences can be minimized by prehybridization of the probe or array with vast excess of unlabeled copies of the repetitive sequences.

Alternative methods have been described for making microarrays of very short DNA sequences, involving photolithography (Pease et al. 1994) or physical masking (Maskos and Smith 1994, 1992) methods. These in situ synthesis methods are inherently limited to low complexity arrays consisting of oligonucleotides. For complex-probe hybridizations, the specificity of

hybridization is improved by using DNA fragments substantially longer than oligonucleotides. Moreover, the in situ synthesis approaches to array fabrication depend on prior knowledge of the sequence to be recognized by each array element. The approach described here makes microarrays by transferring tiny volumes of DNA samples from microwell storage plates to a solid substrate. Thus, nucleic acids (or other molecules) of virtually any length or any origin can be arrayed, and knowledge of their sequences is not required.

The arrays used in these experiments do not represent the maximal achievable density of elements. We have found that the spacing between the spots can be decreased by shrinking the contact area of the printing tip and by increasing the hydrophobicity of the glass surface. Microarrays with 100- μm feature size have been tested successfully in pilot experiments (data not shown). Assuming the projected availability of the appropriate physically mapped human genomic clones (Hudson et al. 1995), arrays at 100- μm spacing would allow for 10,000 discrete intervals of the human genome to be represented in a 1- cm^2 array. Such an array could be used for mapping at a resolution of <0.5 Mb. Experiments are in progress to explore the feasibility of such arrays.

Our initial motivation for developing these microarrays arose from the need for abundant and inexpensive genomic arrays for genomic mismatch scanning (GMS) (Nelson et al. 1993), a method of genetic linkage analysis based on identification of the regions of "identity by descent" between affected relative pairs using a single complex-probe hybridization to an array of genomic clones. Experiments using these arrays to map quantitative trait loci in yeast by GMS are currently in progress (J. deRisi, D. Lashkari, L. Penland, L. McAllister, J. McCusker, R. Davis, and P.O. Brown, unpubl.).

Microarrays of cDNA clones, prepared using the system described here, have been used for quantitative monitoring of gene expression patterns in *Arabidopsis* (Schena et al. 1995), *S. cerevisiae* (D. Lashkari, J. deRisi, L. Penland, P.O. Brown, and R. Davis, unpubl.), and human tissues (J. deRisi, M. Bittner, P. Meltzer, L. Penland, J. Trent, and P.O. Brown, unpubl.). We anticipate that DNA microarrays of the kind described here will be useful in additional applications for which conventional dot blots, high-density gridded arrays on porous membranes, or FISH are currently used. These potential applica-

tions include comparative genomic hybridization (Kallioniemi et al. 1992), sequencing by hybridization (Drmanac et al. 1993), physical mapping of cloned or amplified sequences (Billings et al. 1991), and economical distribution of reagents for integrated genetic and physical mapping based on a common set of arrayed clones (Zehetner and Lehrach 1994).

METHODS

Amplification of Target DNA Elements

The array elements were prepared from physically mapped λ clones (Riles et al. 1993). The λ clones were amplified using randomly primed polymerase chain reaction (PCR) based on published and unpublished protocols (Bohlander et al. 1992; S. Nelson, unpubl.). The phage lysates were amplified in a 10- μ l PCR reaction using 5 μ M final concentration of primer A (GCTATCTTCAAGATCANNNNNN), 200 μ M dNTPs, and 1 unit of *Taq* polymerase. Round A consisted of five cycles at 94°C for 1 min, 25°C for 1.5 min, 25–72°C over 7 min, and 72°C for 3 min using *Taq* polymerase (BMB). For round B, the reaction volume was brought up to 100 μ l for a final concentration of 2 μ M of primer B (GCTATCTTCAAGATCA), 200 μ M dNTPs, and 4 units of *Taq* polymerase. Round B consisted of 30 cycles of 94°C for 1 min, 56°C for 2 min, and 72°C for 3 min. The amplification was performed in 96-well plates using crude phage lysates as the templates, resulting in an amplification of both the 35-kb λ vector and the 5-kb to 15-kb yeast insert sequences as a distribution of PCR products between 250 bp and 1500 bp in length.

The PCR products were purified and transferred into TE (10 mM Tris, 1 mM EDTA at pH 8.0) buffer using Sephadex G50 gel filtration (Pharmacia) and evaporated to dryness at room temperature overnight. Each of the 864 am-

plified λ clones was rehydrated in 15 μ l of 3 \times SSC (20 \times SSC = 3 M NaCl, 0.3 M Na₃ citrate) in preparation for spotting onto the glass under normal room temperature conditions.

Preparation of DNA Microarrays

The microarrays were fabricated on poly-L-lysine coated microscope slides (Sigma). A custom-built arraying machine, consisting of four tweezer-like printing tips mounted 9 mm apart on a computer-controlled robotic stage (Shalon 1996), loaded 1 μ l of the concentrated PCR product directly from corresponding clusters of four wells of 96-well storage plates and deposited ~5 nl of each sample onto each of 40 slides. Surface tension loaded the sample into the printing tip directly from the microwell plate and held the sample in the tip during the printing operation. Printing was achieved by lightly tapping the tip against the glass surface. The open-capillary design allowed for rapid rinsing and drying of the tips between samples. Figure 3 shows the layout of the arraying machine. Figure 4 shows a detailed view of the four printing tips and the staggered printing pattern on the microscope slides. Adjacent samples were spotted 380 μ m apart on the slides. After each set of four samples was printed onto 40 slides, the printing tips were rinsed with a jet of water for 2 sec and then dried by lowering the tips onto a sponge for 2 sec. The process was repeated for all 864 samples and eight control spots.

After the spotting operation was complete, the slides were rehydrated in a humid chamber at room temperature for 2 hr, baked in an 80°C vacuum oven for 2 hr, then rinsed in 0.1% sodium dodecyl sulfate (SDS) to remove unadsorbed DNA. To reduce nonspecific adsorption of the labeled hybridization probe to the poly-L-lysine coated glass surface, the slides were treated with succinic anhydride. One gram of succinic anhydride was dissolved in 100 ml of 1-methyl-2-pyrrolidinone and then 100 ml of 0.2 M boric acid (pH 8.0) was added. The arrays were soaked in this solution for 10 min and then rinsed in distilled water four times for 5 min each. Immediately before use, the arrayed DNA elements were denatured by placing the slide in distilled water at 90°C for 2 min.

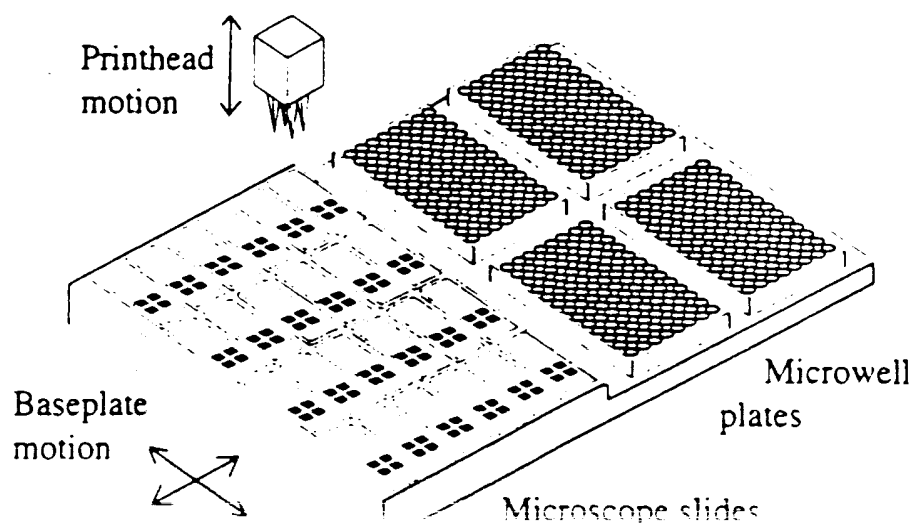


Figure 3. Layout of the arraying machine. All motions are under computer control. For more details of the arraying machine, see web page <http://cmgm.stanford.edu/pbrown>.

Amplification and Labeling of Hybridization Probe

The 16 chromosomes of *Saccharomyces cerevisiae* were separated using a contour-clamped homogeneous electric field (CHEF) agarose gel apparatus (Bio-Rad) (Chu et al. 1986). The 6 largest chromosomes were isolated in one gel slice, and the 10 smallest chromosomes were isolated in a second gel slice. The DNA from each slice was recovered using a gel extraction kit

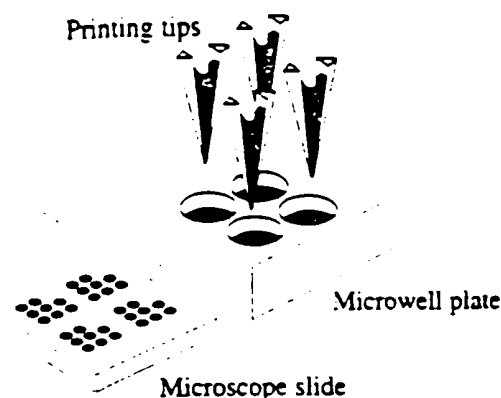


Figure 4 A close-up view of the four open-capillary printing tips. The tips are 9 mm apart and fit into four adjacent wells of a standard microwell plate and print arrays in a staggered fashion on microscope slides. For more details of the printing tips, see web page <http://cmgm.stanford.edu/pbrown>.

(Qiagen) and randomly amplified in a manner similar to that used in amplifying the target λ clones (Grothues et al. 1993). The main difference between this amplification procedure and the one used for the λ array elements is a filtration step between rounds A and B to remove primer-dimers and the use of a random 9-mer 3' end on primer A. Following amplification, 2.5 μ g of each of the amplified chromosome pools were separately random-primer labeled using Klenow polymerase (Amersham) with a lissamine-conjugated nucleotide analog (DuPont NEN) for the pool containing the 6 largest chromosomes and with a fluorescein-conjugated nucleotide analog (BMB) for the pool containing the smallest 10 chromosomes. The two fluorescent-labeled pools were mixed and concentrated using an ultrafiltration device (Amicon).

Hybridization

Five micrograms of the hybridization probe, consisting of both chromosome pools in 7.5 μ l of TE, was denatured in a boiling water bath and then snap-cooled on ice. Concentrated hybridization solution (2.5 μ l) was added to a final concentration of $5 \times$ SSC/0.1% SDS. The entire 10 μ l of probe solution was transferred to the array surface, covered with a coverslip, placed in a custom-built single-slide humidity chamber, and incubated in a 60°C water bath for 12 hr. The custom-built waterproof slide chamber has a cavity just slightly bigger than a microscope slide and was kept at 100% humidity internally by the addition of 2 μ l of water in a corner of the chamber. The slide was rinsed in $5 \times$ SSC/0.1% SDS for 5 min and then in $0.2 \times$ SSC/0.1% SDS for 5 min. All rinses were at room temperature. The array was then air dried, and a drop of antifade (Molecular Probes) was applied to the array under a 24-mm \times 30-mm coverslip in preparation for scanning.

Detection and Analysis

A custom-built laser scanner was used to detect the two-

color fluorescence hybridization signals from 1.8-cm \times 1.8-cm arrays at 20- μ m resolution. The glass substrate slide was mounted on a computer-controlled, two-axis translation stage (PM-500, Newport, Irvine, CA) that scanned the array over an upward-facing microscope objective (20 \times , 0.75NA Fluor, Nikon, Melville, NY) in a bidirectional raster pattern. A water-cooled Argon/Krypton laser (Innova 70 Spectrum, Coherent, Palo Alto, CA), operated in multiline mode, allowed for simultaneous specimen illumination at 488.0 nm and 568.2 nm. These two lines were isolated by a 488/568 dual-band excitation filter (Chroma Technology, Brattleboro, VT). An epifluorescence configuration with a dual-band 488/568 primary beam splitter (Chroma) excited both fluorophores simultaneously and directed fluorescence emissions toward the two-channel detector. Emissions were split by a secondary dichroic mirror with a 565 transition wavelength onto two multialkali cathode photomultiplier tubes (PMT; R928, Hamamatsu, Bridgewater, NJ), one with an HQ535/50 bandpass barrier filter and the other with a D630/60 bandpass barrier filter (Chroma). Preamplified PMT signals were read into a personal computer using a 12-bit analog-to-digital conversion board (RTI-834, Analog Devices, Norwood, MA), displayed in a graphics window, and stored to disk for further rendering and analysis. The back aperture of the 20 \times objective was deliberately underfilled by the illuminating laser beam to produce a large-diameter illuminating spot at the specimen (5- μ m to 10- μ m half-width). Stage scanning velocity was 100 mm/sec, and PMT signals were digitized at 100 μ sec intervals. Two successive readings were summed for each pixel, such that pixel spacing in the final image was 20 μ m. Beam power at the specimen was \sim 5 mW for each of the two lines.

The scanned image was despeckled using a graphics program (Hijaak Graphics Suite) and then analyzed using a custom image gridding program that created a spreadsheet of the average red and green hybridization intensities for each spot. The red and green hybridization intensities were corrected for optical cross talk between the fluorescein and lissamine channels, using experimentally determined coefficients.

ACKNOWLEDGMENTS

This research was supported by grant HG00450 from the National Institutes of Health-National Center for Human Genome Research, a National Science Foundation graduate fellowship to D.S., and by the Howard Hughes Medical Institute. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute. We thank John Mulligan and John McCusker for help in preparing and amplifying the λ clones used in the arrays, Ren Xin Xia for writing the scanner control software and the image gridding and automatic karyotyping programs, Jeff van Ness at Darwin Molecular Corporation for suggesting the use of succinic anhydride, Stan Nelson, Linda McAllister, Joe deRisi, and Lolita Penland for helpful suggestions in the course of this work, and Joe deRisi and Linda McAllister for helpful comments on the manuscript.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

- GENOME RESEARCH ▲▲▲

Discovery and analysis of inflammatory disease-related genes using cDNA microarrays

(inflammation/human genome analysis/gene discovery)

RENU A. HELLER*†, MARK SCHENA*, ANDREW CHAI*, DARI SHALON‡, TOD BEDILION‡, JAMES GILMORE‡, DAVID E. WOOLLEY§, AND RONALD W. DAVIS*

*Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305; ‡SynGene, Palo Alto, CA 94306; and §Department of Medicine, Manchester Royal Infirmary, Manchester, United Kingdom

Contributed by Ronald W. Davis, December 27, 1996

ABSTRACT cDNA microarray technology is used to profile complex diseases and discover novel disease-related genes. In inflammatory disease such as rheumatoid arthritis, expression patterns of diverse cell types contribute to the pathology. We have monitored gene expression in this disease state with a microarray of selected human genes of probable significance in inflammation as well as with genes expressed in peripheral human blood cells. Messenger RNA from cultured macrophages, chondrocyte cell lines, primary chondrocytes, and synovial cells provided expression profiles for the selected cytokines, chemokines, DNA binding proteins, and matrix-degrading metalloproteinases. Comparisons between tissue samples of rheumatoid arthritis and inflammatory bowel disease verified the involvement of many genes and revealed novel participation of the cytokine interleukin 3, chemokine Gro α and the metalloproteinase matrix metallo-elastase in both diseases. From the peripheral blood library, tissue inhibitor of metalloproteinase 1, ferritin light chain, and manganese superoxide dismutase genes were identified as expressed differentially in rheumatoid arthritis compared with inflammatory bowel disease. These results successfully demonstrate the use of the cDNA microarray system as a general approach for dissecting human diseases.

The recently described cDNA microarray or DNA-chip technology allows expression monitoring of hundreds and thousands of genes simultaneously and provides a format for identifying genes as well as changes in their activity (1, 2). Using this technology, two-color fluorescence patterns of differential gene expression in the root versus the shoot tissue of *Arabidopsis* were obtained in a specific array of 48 genes (1). In another study using a 1000 gene array from a human peripheral blood library, novel genes expressed by T cells were identified upon heat shock and protein kinase C activation (3).

The technology uses cDNA sequences or cDNA inserts of a library for PCR amplification that are arrayed on a glass slide with high speed robotics at a density of 1000 cDNA sequences per cm². These microarrays serve as gene targets for hybridization to cDNA probes prepared from RNA samples of cells or tissues. A two-color fluorescence labeling technique is used in the preparation of the cDNA probes such that a simultaneous hybridization but separate detection of signals provides the comparative analysis and the relative abundance of specific genes expressed (1, 2). Microarrays can be constructed from specific cDNA clones of interest, a cDNA library, or a select number of open reading frames from a genome sequencing database to allow a large-scale functional analysis of expressed sequences.

Because of the wide spectrum of genes and endogenous mediators involved, the microarray technology is well suited for analyzing chronic diseases. In rheumatoid arthritis (RA), inflammation of the joint is caused by the gene products of many different cell types present in the synovium and cartilage tissues plus those infiltrating from the circulating blood. The autoimmune and inflammatory nature of the disease is a cumulative result of genetic susceptibility factors and multiple responses, paracrine and autocrine in nature, from macrophages, T cells, plasma cells, neutrophils, synovial fibroblasts, chondrocytes, etc. Growth factors, inflammatory cytokines (4), and the chemokines (5) are the important mediators of this inflammatory process. The ensuing destruction of the cartilage and bone by the invading synovial tissue includes the actions of prostaglandins and leukotrienes (6), and the matrix-degrading metalloproteinases (MMPs). The MMPs are an important class of Zn-dependent metallo-endoproteases that can collectively degrade the proteoglycan and collagen components of the connective tissue matrix (7).

This paper presents a study in which the involvement of select classes of molecules in RA was examined. Also investigated were 1000 human genes randomly selected from a peripheral human blood cell library. Their differential and quantitative expression analysis in cells of the joint tissue, in diseased RA tissue and in inflammatory bowel disease (IBD) tissues was conducted to demonstrate the utility of the microarray method to analyze complex diseases by their pattern of gene expression. Such a survey provides insight not only into the underlying cause of the pathology, but also provides the opportunity to selectively target genes for disease intervention by appropriate drug development and gene therapies.

METHODS

Microarray Design, Development, and Preparation. Two approaches for the fabrication of cDNA microarrays were used in this study. In the first approach, known human genes of probable significance in RA were identified. Regions of the clones, preferably 1 kb in length, were selected by their proximity to the 3' end of the cDNA and for areas of least identity to related and repetitive sequences. Primers were synthesized to amplify the target regions by standard PCR protocols (3). Products were

Abbreviations: RA, rheumatoid arthritis; MMP, matrix-degrading metalloproteinase; IBD, inflammatory bowel disease; LPS, lipopolysaccharide; PMA, phorbol 12-myristate 13-acetate; TNF- α , tumor necrosis factor α ; IL, interleukin; TGF- β , transforming growth factor β ; GCSF, granulocyte colony-stimulating factor; MIP, macrophage inflammatory protein; MIF, migration inhibitory factor; HME, human matrix metallo-elastase; RANTES, regulated upon activation, normal T cell expressed and secreted; Gel, gelatinase; VCAM, vascular cell adhesion molecule-1; E-118, converting enzyme; PMMP, putative metalloproteinase; MnSOD, manganese superoxide dismutase; TIMP, tissue inhibitor of metalloproteinase; MCP, macrophage chemotactic protein.

To whom reprint requests should be sent at the present address: Roche Bioscience, S3-1, 3401 Hillview Avenue, Palo Alto, CA 94304.

One high-resolution color of this article was delayed in part by page number 2150. The authors are grateful to the referees for their comments and for their assistance in the preparation of this manuscript.

Copyright © 1997 by The National Academy of Sciences of the USA. 0897-4224/97/942150-06\$05.00/0
PNAS is available online at <http://www.pnas.org>.

verified by gel electrophoresis and purified with Qiaquick 96-well purification kit (Qiagen, Chatsworth, CA), lyophilized (Savant), and resuspended in 5 μ l of 3 \times standard saline citrate (SSC) buffer for arraying. In the second approach, the microarray containing the 1056 human genes from the peripheral blood lymphocyte library was prepared as described (3).

Tissue Specimens. Rheumatoid synovial tissue was obtained from patients with late stage classic RA undergoing remedial synovectomy or arthroplasty of the knee. Synovial tissue was separated from any associated connective tissue or fat. One gram of each synovial specimen was subjected to RNA extraction within 40 min of surgical excision, or explants were cultured in serum-free medium to examine any changes under *in vitro* conditions. For IBD, specimens of macroscopically inflamed lower intestinal mucosa were obtained from patients with Crohn disease undergoing remedial surgery. The hypertrophied mucosal tissue was separated from underlying connective tissue and extracted for RNA.

Cultured Cells. The Mono Mac-6 (MM6) monocytic cells (8) were grown in RPMI medium. Human chondrosarcoma SW1353 cells, primary human chondrocytes, and synoviocytes (9, 10) were cultured in DMEM; all culture media were supplemented with 10% fetal bovine serum, 100 μ g/ml streptomycin, and 500 units/ml penicillin. Treatment of cells with lipopolysaccharide (LPS) endotoxin at 30 ng/ml, phorbol 12-myristate 13-acetate (PMA) at 50 ng/ml, tumor necrosis factor α (TNF- α) at 50 ng/ml, interleukin (IL)-1 β at 30 ng/ml, or transforming growth factor- β (TGF- β) at 100 ng/ml is described in the figure legends.

Fluorescent Probe, Hybridization, and Scanning. Isolation of mRNA, probe preparation, and quantitation with *Arabidopsis* control mRNAs was essentially as described (3) except for the following minor modification. Following the reverse transcriptase step, the appropriate Cy3- and Cy5-labeled samples were pooled; mRNA degraded by heating the sample to 65°C for 10 min with the addition of 5 μ l of 0.5M NaOH plus 0.5 ml of 10 mM EDTA. The pooled cDNA was purified from unincorporated nucleotides by gel filtration in Centri-spin columns (Princeton Separations, Adelphia, NJ). Samples were lyophilized and dissolved in 6 μ l of hybridization buffer (5 \times SSC plus 0.2% SDS). Hybridizations, washes, scanning, quantitation procedures, and pseudocolor representations of fluorescent images have been described (3). Scans for the two fluorescent probes were normalized either to the fluorescence intensity of *Arabidopsis* mRNAs spiked into the labeling reactions (see Figs. 2–4) or to the signal intensity of β -actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH; see Fig. 5).

RESULTS

Ninety-Six-Genes Microarray Design. The actions of cytokines, growth factors, chemokines, transcription factors, MMPs, prostaglandins, and leukotrienes are well recognized in inflammatory disease, particularly RA (11–14). Fig. 1 displays the selected genes for this study and also includes control cDNAs of housekeeping genes such as β -actin and GAPDH and genes from *Arabidopsis* for signal normalization and quantitation (row A, columns 1–12).

Defining Microarray Assay Conditions. Different lengths and concentrations of target DNA were tested by arraying PCR-

	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	BLANK	HAT1 HAT1	HAT1 HAT1	HAT4 HAT4	HAT4 HAT4	HAT22 HAT22	HAT22 HAT22	YES23 YES23	YES23 YES23	BACTIN β -actin	G3PDH G3PDH
B	IL1A IL-1 α	IL1B IL-1 β	IL1RA IL-1RA	IL2 IL-2	IL3 IL-3	IL4 IL-4	IL6 IL-6	IL6R IL-6R	IL7 IL-7	CFOS c-fos	CJUN c-jun	RFRA1 Rat Fra-1
C	IL8 IL-8	IL9 IL-9	IL10 IL-10	ICE ICE	IFNG IFN γ	GCSF G-CSF	MCSF M-CSF	GMCSF GM-CSF	TNFB.1 TNF β	CREL c-rel	NFKB50 NF- κ Bp50	NFKB65.1 NF- κ Bp65
D	TNFA.1 TNF α	TNFA.2 TNF α	TNFA.3 TNF α	TNFA.4 TNF α	TNFA.5 TNF α	TNFR1.1 TNFR1	TNFR1.2 TNFR1	TNFR1.1 TNFR1	TNFR1.2 TNFR1	NFKB65.2 NF- κ Bp65	IKB I κ B	CREB2 CREB2
E	STR1 Strom-1	STR2.3 Strom-2	STR3 Strom-3	COL1 Coll-1	COL1.3 Coll-1.3	COL2.1 Coll-2	COL2.2 Coll-2	COL3 Coll-3	COX1 Cox-1	COX2 Cox-2	12LO 12-L0	15LO 15-LO
F	GELA.1 Gel-A	GELB Gel-B	HME Elastase	MTMMP MT-MMP	PUMP1 Matrilysin	TIMP1 TIMP-1	TIMP2 TIMP-2	TIMP3 TIMP-3	ICAM1 ICAM-1	VCAM VCAM	5LO.1 5-LO	CPLA2.2 cPLA2
G	EGF EGF	FGFA FGF acidic	FGFB FGF basic	IGF IGF-I	IGFI IGF-II	TGFA TGF α	TGFB TGF β	PDGFB PDGF β	CALCTN Calcitonin	GH GH-1	GRO GRO1 α	GCR GR α
H	MCP1.1 MCP-1	MCP1.1 MCP-1	MIP1A MIP-1 α	MIP1B MIP-1 β	MIF MIF	RANTES RANTES	INOS iNOS	LDLR LDLR	ALU.1 IL-10	ALU.2 TNFRp70	ALU.3 IL-10	POLYA LDLR

A. thaliana controls

Human controls

Cytokines and related genes

Transcription factors and related genes

MMP's and related genes

Chemokines

Growth factors and related genes

Other genes

FIG. 1. Ninety-six-element microarray design. The target element name and the corresponding gene are shown in the layout. Some genes have more than one target element to guarantee specificity of signal. For TNF the targets represent decreasing lengths of 1, 0.8, 0.6, 0.4, and 0.2 kb from left to right.

ically these cells, when triggered by an immunogen, produce the proinflammatory cytokines TNF and IL-1. We have used the monocyte cell line MM6 and monitored changes in gene expression upon activation with LPS endotoxin, a component of Gram-negative bacterial membranes, and PMA, which augments the action of LPS on TNF production (15). RNA was isolated at different times after induction and used for cDNA probe preparation. From this time course it was clear that TNF expression was induced within 15 min of treatment, reached maximum levels in 1 hr, remained high until 4 hr and subsequently declined (Fig. 24). Many other cytokine genes were also transiently activated, such as IL-1 α and - β , IL-6, and granulocyte colony-stimulating factor (GCSF). Prominent chemokines activated were IL-8, macrophage inflammatory protein (MIP)-1 β , more so than MIP-1 α , and Gro α or melanoma growth stimulatory factor. Migration inhibitory factor (MIF) expressed in the uninduced state declined in LPS-activated cells. Of the immediate early genes, the noticeable ones were *c-fos*, *fra-1*, *c-jun*, NF- κ Bp50, and I κ B, with *c-rel* expression observed even in the uninduced state (Fig. 2B). These expression patterns are consistent with reported patterns of activation of certain LPS- and PMA-induced genes (12). Demonstrated here is the unique ability of this system to allow parallel visualization of a large number of gene activities over a period of time.

SW1353 cells is a line derived from malignant tumors of the cartilage and behaves much like the chondrocytes upon stimulation with TNF and IL-1 in the expression of MMPs (9). In addition to confirming our earlier observations with Northern blots on Strom-1, Col-1, and Col-3 expression (9), gelatinase (Gel) A, putative metalloproteinase (PUMP)-1 membrane-

type matrix metalloproteinase, tissue inhibitors of matrix metalloproteinases or tissue inhibitor of metalloproteinase 1 (TIMP-1), -2, and -3 were also expressed by these cells together with the human matrix metallo-elastase (HME; Fig. 34). HME induction was estimated to be ~50-fold and was greater than any of the other MMPs examined (Fig. 3B). This result was unexpected because HME is reportedly expressed only by alveolar macrophage and placental cells (16). Expression of the cytokines and chemokines, IL-6, IL-8, MIF, and MIP-1 β was also noted. A variety of other genes, including certain transcription factors, were also up-regulated (Fig. 3), but the overall time-dependent expression of genes in the SW1353 cells was qualitatively distinct from the MM6 cells.

Quantitation of differential gene expression (Figs. 2B and 3B) was achieved with the simultaneous hybridization of Cy3-labeled cDNA from untreated cells and Cy5-labeled cDNA from treated samples. The estimated increases in expression from these microarrays for a select number of genes including IL-1 β , IL-8, MIP-1 β , TNF, HME, Col-1, Col-3, Strom-1, and Strom-2 were compared with data collected from dot blot analysis. Results (not shown) were in close agreement and confirmed our earlier observations on the use of the microarray method for the quantitation of gene expression (3).

Expression Profiles in Primary Chondrocytes and Synovialocytes of Human RA Tissue. Given the sensitivity and the specificity of this method, expression profiles of primary synovialocytes and chondrocytes from diseased tissue were examined. Without prior exposure to inducing agents, low level expression of *c-jun*, GCSF, IL-3, TNF- β , MIF, and RANTES (regulated upon activation, normal T cell expressed and secreted) was seen as well as expression of MMPs. GelA, Strom-1, Col-1, and the three TIMPs. In this case, Col-2 hybridization was considered to be nonspecific because the second Col-2 target taken from the 3' end of the gene gave no

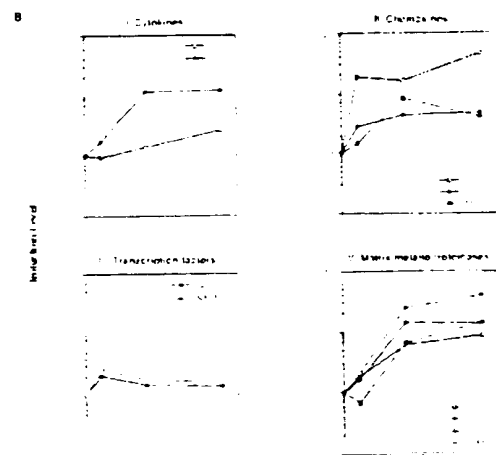
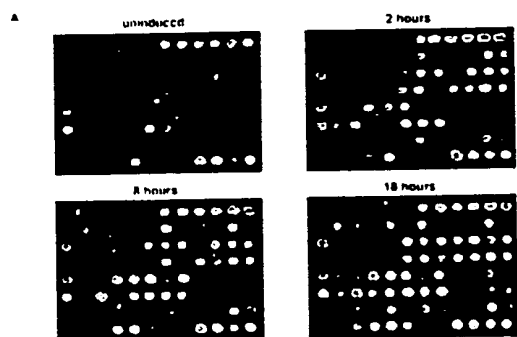


Fig. 2. (A) Fluorescent microarray images showing gene expression levels in MM6 cells at 0, 2, 8, and 18 hours after induction with LPS and PMA. (B) Relative levels of selected genes at different time points compared with time zero.

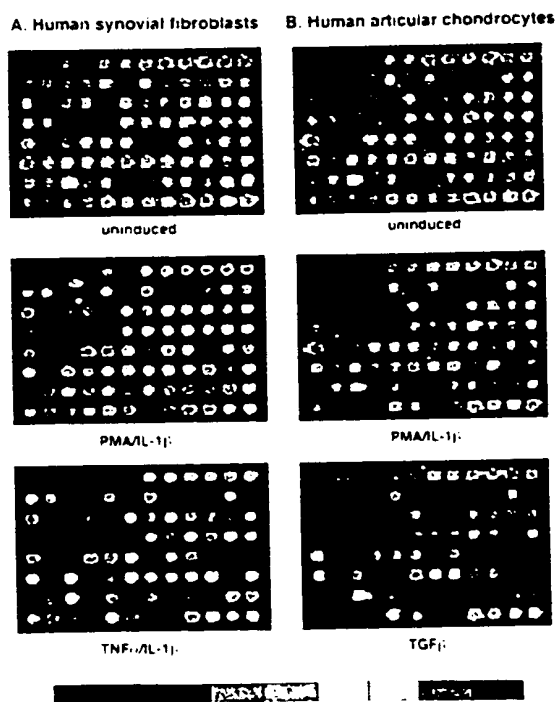


Fig. 3. (A) Fluorescent microarray images showing gene expression levels in human synovial fibroblasts. (B) Fluorescent microarray images showing gene expression levels in human articular chondrocytes. The color bars provide a comparative calibration scale between arrays and are derived from the *Arabidopsis* mRNA samples that are introduced in equal amounts during probe preparation.

signal. Treatment more so with PMA and IL-1, than TNF and IL-1, produced a dramatic up-regulation in expression of several genes in both of these primary cell types. These genes are as follows: the cytokine IL-6, the chemokines IL-8 and Gro-1 α , and the MMPs; Strom-1, Col-1, Col-3, and HME; and the adhesion molecule, vascular cell adhesion molecule 1 (VCAM-1). The surprise again is HME expression in these primary cells, for reasons discussed above. From these results, the expression profiles of synoviocytes and the chondrocytes appear very similar; the differences are more quantitative than qualitative. Treatment of the primary chondrocytes with the anabolic growth factor TGF- β had an interesting profile in that it produced a remarkable down-regulation of genes expressed in both the untreated and induced state (Fig. 4).

Given the demonstrated effectiveness of this technology, a comparative analysis of two different inflammatory disease states was conducted with probes made from RA tissue and IBD samples. RA samples were from late stage rheumatoid synovial tissue, and IBD specimens were obtained from inflamed lower intestinal mucosa of patients with Crohn disease. With both the 96-element known gene microarray and the 1000-gene microarray of cDNAs selected from a peripheral human blood cell library (3), distinct differences in gene expression patterns were evident. On the 96-gene array, RA tissue samples from different affected individuals gave similar profiles (data not shown) as did different samples from the same individual (Fig. 5). These patterns were notably similar to those observed with primary synoviocytes and chondrocytes (Fig. 4). Included in the list of prominently up-regulated genes are IL-6, the MMPs Strom-1, Col-1, GelA, HME, and in

certain samples PUMP, TIMPs, particularly TIMP-1 and TIMP-3, and the adhesion molecule VCAM. Discernible levels of macrophage chemotactic protein 1 (MCP-1), MIF and RANTES were also noted. IBD samples were in comparison, rather subdued although IL-1 converting enzyme (ICE), TIMP-1, and MIF were notable in all the three different IBD samples examined here. In IBD-A, one of three individual samples, ICE, VCAM, Gro-1, and MMP expression was more pronounced than in the others.

We also made use of a peripheral blood cDNA library (3) to identify genes expressed by lymphocytes infiltrating the inflamed tissues from the circulating blood. With the 1046-element array of randomly selected cDNAs from this library, probes made from RA and IBD samples showed hybridizations to a large number of genes. Of these, many were common between the two disease tissues while others were differentially expressed (data not shown). A complete survey of these genes was beyond the scope of this study, but for this report we picked three genes that were up-regulated in the RA tissue relative to IBD. These cDNAs were sequenced and identified by comparison to the GenBank database. They are TIMP-1, apoferritin light chain, and manganese superoxide dismutase (MnSOD). Differential expression of MnSOD was only observed in samples of RA tissue explants maintained in growth medium without serum for anywhere between 2 to 16 hr. These results also indicate that the expression profile of genes can be altered when explants are transferred to culture conditions.

DISCUSSION

The speed, ease, and feasibility of simultaneously monitoring differential expression of hundreds of genes with the cDNA microarray based system (1-3) is demonstrated here in the analysis of a complex disease such as RA. Many different cell types in the RA tissue; macrophages, lymphocytes, plasma cells, neutrophils, synoviocytes, chondrocytes, etc. are known to contribute to the development of the disease with the expression of gene products known to be proinflammatory. They include the cytokines, chemokines, growth factors, MMPs, eicosanoids, and others (7, 11-14), and the design of the 96-element known gene microarray was based on this knowledge and depended on the availability of the genes. The technology was validated by confirming earlier observations on the expression of TNF by the monocyte cell line MM6, and of Col-1 and Col-3 expression in the chondrosarcoma cells and articular chondrocytes (9, 12). In our time-dependent survey the chronological order of gene activities in and between gene families was compared and the results have provided unprecedented profiles of the cytokines (TNF, IL-1, IL-6, GCSF, and MIF), chemokines (MIP-1 α , MIP-1 β , IL-8, and Gro-1), certain transcription factors, and the matrix metalloproteinases (GelA, Strom-1, Col-1, Col-3, HME) in the macrophage cell line MM6 and in the SW1353 chondrosarcoma cells.

Earlier reports of cytokine production in the diseased state had established a model in which TNF is a major participant in RA. Its expression reportedly preceded that of the other cytokines and effector molecules (4). Our results strongly support these results as demonstrated in the time course of the MM6 cells where TNF induction preceded that of IL-1 α and IL-1 β followed by IL-6 and GCSF. These expression profiles demonstrate the utility of the microarrays in determining the hierarchy of signaling events.

In the SW1353 chondrosarcoma cells, all the known MMPs and TIMPs were examined simultaneously. HME expression was discovered, which previously had been observed in only the stromal cells and alveolar macrophages of smoker's lungs and in placental tissue. Its presence in cells of the RA tissue is meaningful because its activity can cause significant destruction of articular matrix. Expression profiles of synovial fibroblasts and articular chondrocytes were remarkably similar and not too different from the SW1353 cells, indicating that the fibroblast and the chondrocyte can play equally aggressive roles in joint erosion. Prominent genes expressed were

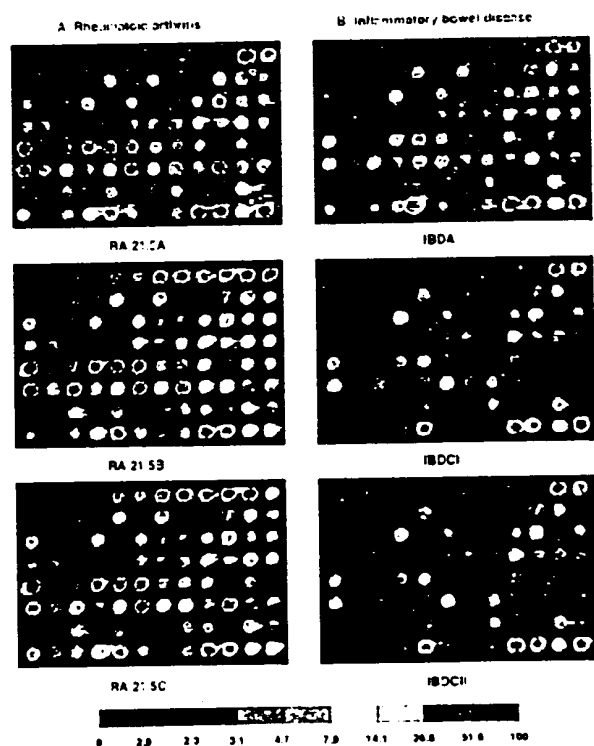


Fig. 5. Expression profiles of RA tissue (A) and IBD tissue (B). mRNA from RA tissue samples obtained from the same individual was isolated directly after excision (RA 21 SA) or maintained in culture for 2 hr before isolation (RA 21 SB and RA 21 SC). For the IBD tissue samples, only one individual (data not shown) was remarkably similar to the ones shown here (IBD A and IBD C1) from mRNA samples prepared directly after surgery from two separate individuals. For the IBD-C2 probe, the tissue sample was cultured in medium without serum for 2 hr before mRNA preparation.

the MMPs, but chemokines and cytokines were also produced by these cells. The effect of the anabolic growth factor TGF- β was profoundly evident in demonstrating the down regulation of these catabolic activities.

RA tissue samples undeniably reflected profiles similar to the cell types examined. Active genes observed were IL-3, IL-6, ICE, the MMPs including HME and TIMPs, chemokines IL-8, Gro α , MIP, MIF, and RANTES, and the adhesion molecule VCAM. Of the growth factors, fibroblast growth factor β was observed most frequently. In comparison, the expression patterns in the other inflammatory state (i.e., IBD) were not as marked as in the RA samples, at least as obtained from the tissue samples selected for this study.

As an alternative approach, the 1046 cDNA microarray of randomly selected genes from a lymphocyte library was used to identify genes expressed in RA tissue (3). Many genes on this array hybridized with probes made from both RA and IBD tissue samples. The results are not surprising because inflammatory tissue is abundantly supplied with cell types infiltrating from the circulating blood, made apparent also by the high levels of chemokine expression in RA tissue. Because of the magnitude of the effort required to identify all the hybridized genes, we have for this report chosen to describe only three differentially expressed genes mainly to verify this method of analysis.

Of the large number of genes observed here, a fair number were already known as active participants in inflammatory disease. These are TNF, IL-1, IL-6, IL-8, GCSF, RANTES, and VCAM. The novel participants not previously reported are HME, IL-3, ICE, and Gro α . With our discovery of HME expression in RA, this gene becomes a target for drug intervention. ICE is a cysteine protease well known for its IL-1 β processing activity (18), and recognized for its role in apoptotic cell death (19). Its expression in RA tissue is intriguing. IL-3 is recognized for its growth-promoting activity in hematopoietic cell lineages, is a product of activated T cells (20), and its expression in synovocytes and chondrocytes of RA tissue is a novel observation.

Like IL-8, Gro α , is a C-X-C subgroup chemokine and is a potent neutrophil and basophil chemoattractant. It down-regulates the expression of types I and III interstitial collagens (21, 22) and is seen here produced by the MM6 cells, in primary synovocytes, and in RA tissue. With the presence of RANTES, MCP, and MIP-1 β , the C-C chemokines (23) migration and infiltration of monocytes, particularly T cells, into the tissue is also enhanced (5) and aid in the trafficking and recruitment of leukocytes into the RA tissue. Their activation, phagocytosis, degranulation, and respiratory bursts could be responsible for the induction of MnSOD in RA. MnSOD is also induced by TNF and IL-1 and serves a protective function against oxidative damage. The induction of the ferritin light chain encoding gene in this tissue may be for reasons similar to those for MnSOD. Ferritin is the major intracellular iron storage protein and it is responsive to intracellular oxidative stress and reactive oxygen intermediates generated during inflammation (24, 25). The active expression of TIMP-1 in RA tissue, as detected by the 1000-element array, is no surprise because our results have repeatedly shown TIMP-1 to be expressed in the constitutive and induced states of RA cells and tissues.

The suitability of the cDNA microarray technology for profiling diseases and for identifying disease related genes is well documented here. This technology could provide new

targets for drug development and disease therapies, and in doing so allow for improved treatment of chronic diseases that are challenging because of their complexity.

We would like to thank the following individuals for their help in obtaining reagents or providing cDNA clones to use as templates in target preparation: N. Arai, P. Cannon, D. R. Cohen, T. Curran, V. Dixit, D. A. Geller, G. I. Goldberg, M. Karin, M. Lotz, L. Matrisian, G. Nolan, C. Lopez-Otin, T. Schall, S. Shapiro, I. Verma, and H. Van Wart. Support for R.W.D., M.S., and R.A.H. was provided by the National Institutes of Health (Grants R37HG00198 and HG00205).

1. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* **270**, 467-470.
2. Shalon, D., Smith, S. & Brown, P. O. (1996) *Genome Res.* **6**, 639-645.
3. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. & Davis, R. W. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 10614-10619.
4. Feldmann, M., Brennan, F. M. & Maini, R. N. (1996) *Rheumatoid Arthritis Cell* **85**, 307-310.
5. Schall, T. J. (1994) in *The Cytokine Handbook*, ed. Thomson, A. W. (Academic, New York), 2nd Ed., pp. 410-460.
6. Lotz, M. F., Blanco, J., Von Kempis, J., Dudler, J., Maier, R., Villiger, P. M. & Geng, Y. (1995) *J. Rheumatol.* **22**, Supplement 43, 104-108.
7. Birkedal-Hansen, H., Moore, W. G. I., Bodden, M. K., Windsor, L. J., Birkedal-Hansen, B., DeCarlo, A. & Engler, J. A. (1993) *Crit. Rev. Oral Biol. Med.* **4**, 197-250.
8. Zeigler-Heitbrock, H. W. L., Thiel, E., Futterer, A., Volker, H., Wirtz, A. & Reithmuller, G. (1988) *Int. J. Cancer* **41**, 456-461.
9. Borden, P., Solymar, D., Sucharczuk, A., Lindman, B., Cannon, P. & Heller, R. A. (1996) *J. Biol. Chem.* **271**, 23577-23581.
10. Gadhra, S. J. & Woolley, D. E. (1987) *Rheumatol. Int.* **7**, 13-22.
11. Harris, E. D., Jr. (1990) *New Engl. J. Med.* **322**, 1277-1289.
12. Firestein, G. S. (1996) in *Textbook of Rheumatology*, eds. Kelly, W. N., Harris, E. D., Ruddy, S. & Sledge, C. B. (Saunders, Philadelphia), 5th Ed., pp. 5001-5047.
13. Alvaro-Garcia, J. M., Zvaifler, Nathan J., Brown, C. B., Kaushansky, K. & Firestein, Gary S. (1991) *J. Immunol.* **146**, 3365-3371.
14. Firestein, G. S., Alvaro-Garcia, J. M. & Maki, R. (1990) *J. Immunol.* **144**, 3347-3352.
15. Pradines-Figueres, A. & Raetz, C. R. H. (1992) *J. Biol. Chem.* **267**, 23261-23268.
16. Shapiro, S. D., Kobayashi, D. L. & Ley, T. J. (1993) *J. Biol. Chem.* **268**, 23824-23829.
17. Shipley, M. J., Wesselschmidt, R. L., Kobayashi, D. K., Ley, T. J. & Shapiro, S. D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 3042-3046.
18. Cerretti, D. P., Kozlosky, C. J., Mosley, B., Nelson, N., Van Ness, K., Greenstreet, T. A., March, C. J., Kronheim, S. R., Druck, T., Cannizaro, L. A., Huebner, K. & Black, R. A. (1992) *Science* **256**, 97-100.
19. Miura, M., Zhu, H., Rotello, R., Hartweg, E. A. & Yuan, J. (1993) *Cell* **75**, 653-660.
20. Arai, K., Lee, F., Miyajima, A., Shoichiro, M., Arai, N. & Takashi, Y. (1990) *Annu. Rev. Biochem.* **59**, 783-836.
21. Geiser, T., Dewald, B., Ehrengruber, M. U., Lewis, I. C. & Baggiolini, M. (1993) *J. Biol. Chem.* **268**, 15419-15424.
22. Unemori, E. N., Amento, E. P., Bauer, E. A. & Horuk, R. (1993) *J. Biol. Chem.* **268**, 1338-1342.
23. Robinson, E., Keystone, E. C., Schall, T. J., Gillet, N. & Fish, E. N. (1995) *Clin. Exp. Immunol.* **101**, 398-407.
24. Roeser, H. (1980) in *Iron Metabolism in Biochemistry and Medicine*, eds. Jacobs, A. & Worwood, M. (Academic, New York), Vol. 2, pp. 605-640.
25. Kwak, E. L., Larochelle, D. A., Beaumont, C., Torti, S. V. & Torti, F. M. (1995) *J. Biol. Chem.* **270**, 15285-15293.

1

Molecular Cloning

A LABORATORY MANUAL

SECOND EDITION

J. Sambrook

UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER

E.F. Fritsch

GENETICS INSTITUTE

T. Maniatis

HARVARD UNIVERSITY



Cold Spring Harbor Laboratory Press
1989

Molecular Cloning

A LABORATORY MANUAL
SECOND EDITION

All rights reserved
© 1989 by Cold Spring Harbor Laboratory Press
Printed in the United States of America

9 8 7 6 5 4 3 2 1

Book and cover design by Emily Harste

Cover: The electron micrograph of bacteriophage λ particles stained with uranyl acetate was digitized and assigned false color by computer. (Thomas R. Broker, Louise T. Chow, and James I. Garrels)

Cataloging in Publications data

Sambrook, Joseph

Molecular cloning: a laboratory manual / E.F.

Fritsch, T. Maniatis—2nd ed.

p. cm.

Bibliography: p.

Includes index.

ISBN 0-87969-309-6

1. Molecular cloning—Laboratory manuals. 2. Eukaryotic cells—Laboratory manuals. I. Fritsch, Edward F. II. Maniatis, Thomas III. Title.

QH442.2.M26 1987

574.87'3224—dc19

87-35464

Researchers using the procedures of this manual do so at their own risk. Cold Spring Harbor Laboratory makes no representations or warranties with respect to the material set forth in this manual and has no liability in connection with the use of these materials.

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Cold Spring Harbor Laboratory Press for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of \$0.10 per page is paid directly to CCC, 21 Congress St., Salem MA 01970. [0-87969-309-6/89 \$00 + \$0.10] This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale.

All Cold Spring Harbor Laboratory Press publications may be ordered directly from Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, New York 11724. Phone: 1-800-843-

Analysis of RNA

A number of methods have been developed to quantitate, measure the size of, and map the 5' and 3' termini of specific mRNA molecules in preparations of cellular RNA. These include:

- *Northern hybridization (RNA blotting)*, in which the size and amount of specific mRNA molecules in preparations of total or poly(A)⁺ RNA are determined (Alwine et al. 1977, 1979). The RNA is separated according to size by electrophoresis through a denaturing agarose gel and is then transferred to activated cellulose (Alwine et al. 1977; Seed 1982b), nitrocellulose (Goldberg 1980; Thomas 1980; Seed 1982a), or glass or nylon membranes (Bresser and Gillespie 1983) (see below). The RNA of interest is then located by hybridization with radiolabeled DNA or RNA followed by autoradiography.
- *Dot and slot hybridization*, in which an excess of radiolabeled probe is hybridized to RNA that has been immobilized on a solid support (Kafatos et al. 1979; Thomas 1980; White and Bancroft 1982). Densitometric tracings of the resulting autoradiographs can allow comparative estimates of the amount of the target sequence in various preparations of RNA.
- *Mapping RNA using nuclease S1 or ribonuclease*, in which the precise positions of the 5' and 3' termini of the mRNA and the locations of splice junctions can be rigorously determined (Berk and Sharp 1977; Weaver and Weissmann 1979). Labeled or unlabeled RNA or DNA probes derived from various segments of the genomic DNA are hybridized to mRNA, often under conditions favoring the formation of DNA:RNA hybrids (Casey and Davidson 1977). The products of the hybridization are then digested with nuclease S1 or RNAase under conditions favoring digestion of single-stranded nucleic acids only. Analysis of the digestion products by gel electrophoresis yields important quantitative and qualitative information about the mRNA structure.
- *Primer extension*, in which a small radiolabeled fragment of DNA is hybridized to the mRNA and used as a primer for reverse transcriptase. The resulting product should extend to the extreme 5' terminus of the mRNA, and thus the size of the product reflects the number of nucleotides from the position of the label to the 5' terminus of the mRNA.
- *Solution hybridization*, in which the absolute concentration of the sequence of interest is calculated from the rate of hybridization of a small amount of a specific radioactive probe with a known quantity of purified cellular RNA (see, e.g., Roop et al. 1978; Durnam and Palmiter 1983). Alternatively, an excess of a radiolabeled probe is incubated with a known amount of RNA. The concentration of the RNA of interest can then be estimated from the amount of radioactivity that becomes resistant to nuclease S1 (see, e.g., Favalerio et al. 1980; Beach and Palmiter 1981; Williams et al. 1986).

- *Filter hybridization*, in which purified cellular RNA is end-labeled with ^{32}P and hybridized to a large excess of the homologous DNA that has been immobilized on a solid support (Williams et al. 1986).

Below we describe northern hybridization. Dot and slot hybridization of both crude and purified preparations of RNA are described beginning on page 7.53; nuclease-S1 and RNAase analysis of specific hybrids, beginning on pages 7.58 and 7.71, respectively; and analysis of mRNA by primer extension, beginning on page 7.79.